(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 October 2004 (14.10.2004)

PCT

(10) International Publication Number
WO 2004/087874 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US2004/009202

(22) International Filing Date: 24 March 2004 (24.03.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/458,824 28 March 2003 (28.03.2003) US

(71) Applicants *(for all designated States except US)*: **NUVELO, INC.** [US/US]; 675 Almanor Avenue, Sunnyvale, CA 94085 (US). **DRMANAC, Radoje, T.** [US/US]; 27635 Red Rock Road, Los Altos Hills, CA 94022 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ZHOU, Ping** [US/US]; 7595 Newcastle Drive, Cupertino, CA 95014 (US). **WANG, Jian-Rui** [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). **WANG, Zhi, Wei** [CN/US]; 114 Indiana Avenue, Athens, GA 30605 (US). **HU, Tianhua** [CN/US]; 1638 Roberta Drive, San Mateo, CA 94403 (US).

(74) Agent: **POLIZOTTO, Renee**; Nuvelo, Inc., 675 Almanor Avenue, Sunnyvale, CA 94085 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/087874 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences enclosed by these nucleic acids and uses thereof.

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1.1 CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority to U.S. Provisional Application Serial No. 60/458,824 filed March 28, 2003 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 824. Related subject matter is disclosed in the following applications:

a) U.S. Application Serial No. 10/296,115 (I.A. filing date of December 22, 2000) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 784CIP3A/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US00/35017 filed December 22, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP3A/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/552,317 filed April 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/488,725 filed January 21, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784;

b) U.S. Application No. 10/275,027 (I.A. filing date of January 25, 2001) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 785CIP3/PCT which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/02623 filed January 25, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/491,404 filed January 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785 (now abandoned);

c) U.S. Application Serial No. 10/276,774 (I.A. filing date of February 5, 2001) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 787CIP3/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/03800 filed February 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/560,875 filed April 27, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/496,914 filed February 03, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787 (now abandoned);

d) U.S. Application Serial No. 10/220,366 (I.A. filing date of February 26, 2001) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 788CIP3/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/04927 filed February 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,409 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/515,126 filed February 28, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788 (now abandoned);

e) U.S. Application Serial No. 10/221,279 (I.A. filing date of March 5, 2001) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 789CIP3/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/04941 filed March 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/574,454 filed May 19, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/519,705 filed March 07, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789 (now abandoned);

f) U.S. Application Serial No. 10/450,763 (I.A. filing date of March 30, 2001) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 790CIP3/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/08631 filed March 30, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/649,167 filed August 23, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/540,217 filed March 31, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790 (now abandoned);

g) PCT Application Serial No. PCT/US01/08656 filed April 18, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/770,160 filed

January 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP (now abandoned), which is in turn a continuation-in-part application of U.S. Application Serial No. 09/552,929 filed April 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791 (now abandoned);

h) U.S. Application Serial No. 10/276,817 (I.A. filing date of May 16, 2001) entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 792CIP3/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/14827 filed May 16, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,408 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792;

i) U.S. Application Serial No. 10/461,673 filed June 13, 2003 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 823, which is a continuation-in-part application of 1) U.S. Application Serial No. 10/363,616 (I.A. filing date of August 31, 2001) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 793CIP/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/27093 filed August 31, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 793CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/654,935 filed September 01, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 793; 2) U.S. Application Serial No. 10/380,731 (I.A. filing date of September 10, 2001) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 794CIP/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/26015 filed September 10, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 794CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/659,671 filed September 11, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 794; 3) U.S. Application Serial No. 10/399,103 (I.A. filing date of October 11, 2001) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 795CIP/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US01/27760 filed October 11, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 795CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/687,527 filed October 12, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 795 (now abandoned); 4) PCT Application

Serial No. PCT/US01/42950 filed November 16, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 797CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/714,936 filed November 17, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 797; 5) PCT Application Serial No. PCT/US01/47004 filed November 30, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 799CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/728,952 filed November 30, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 799; 6) PCT Application Serial No. PCT/US02/01222 filed January 29, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 802CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/774,528 filed January 30, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 802; 7) PCT Application Serial No. PCT/US02/05095 filed March 05, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 803CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/799,451 filed March 05, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 803; 8) PCT Application Serial No. PCT/US02/05109 filed March 14, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 804CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/810,173 filed March 15, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 804 (now abandoned); 9) PCT Application Serial No. PCT/US02/22858 filed July 19, 2002 entitled "Novel Nucleic Acids and Secreted Polypeptides," Attorney Docket No. 805A/PCT which claims the benefit of priority to U.S. Provisional Application Serial No. 60/306,971 filed July 21, 2001 entitled "Novel Nucleic Acids and Secreted Polypeptides," Attorney Docket No. 805 (now expired); 10) PCT Application Serial No. PCT/US02/25485 filed August 09, 2002 entitled "Novel Nucleic Acids and Secreted Polypeptides," Attorney Docket No. 806CIP/PCT claims the benefit of priority to U.S. Provsional Application Serial No. 60/311,261 filed August 09, 2001 entitled "Novel Nucleic Acids and Secreted Polypeptides," Attorney Docket No. 806 (now expired); 11) PCT Application Serial No. PCT/US02/29001 filed September 13, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 807ACIP/PCT which claims the benefit of priority to U.S. Provisional Application Serial No. 60/322,511 filed September 13, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 807 (now expired); 12) PCT Application Serial No. PCT/US02/29636 filed September 18, 2002

entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 808ACIP/PCT which claims the benefit of priority to U.S. Provisional Application Serial No. 60/323,349 filed September 18, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 808 (now expired); and 13) PCT Application Serial No. PCT/US02/29964 filed September 19, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 809ACIP/PCT which claims the benefit of priority to U.S. Provisional Application Serial No. 60/323,739 filed September 19, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 809 (now expired);

j) PCT Application Serial No. PCT/US01/02723 filed January 25, 2001 entitled "Novel Fetal Nucleic Acids and Polypeptides," Attorney Docket No. 796/785CIP/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/707,351 filed November 06, 2000 entitled "Novel Fetal Nucleic Acids and Polypeptides," Attorney Docket No. 796 (now abandoned);

k) U.S. Application Serial No. _____ (I.A. filing date of September 24, 2002) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 810CIP/US which is a U.S. National Application filed under 35 U.S.C. 371 of PCT Application Serial No. PCT/US02/30474 filed September 24, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 810CIP/PCT, which in turn claims the priority benefit of U.S. Provisional Application Serial No. 60/324,631 filed September 24, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 810 (now expired);

l) PCT Application Serial No. PCT/US02/39555 filed December 10, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 820/PCT, which in turn claims the priority benefit of U.S. Provisional Application Serial No. 60/339,739 filed December 10, 2001 entitled "Novel Nucleic Acids and Secreted Polypeptides," Attorney Docket No. 811 (now expired), and is a continuation-in-part application of U.S. Application Serial No. 10/128,558 filed April 22, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 812A, which claims the benefit of priority to U.S. Provisional Application Serial No. 60/339,453 filed December 11, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 812 (now expired), and contains related subject matter that is disclosed in U.S. Provisional Application Serial Nos. 60/340,187 filed December 12, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 813 (now expired), 60/365,384 filed March 14, 2002 entitled "Novel Nucleic Acids and Secreted Polypeptides," Attorney Docket No. 814 (now expired), 60/365,091 filed March 14, 2002 entitled "Novel

Nucleic Acids and Polypeptides," Attorney Docket No. 815 (now expired), 60/365,264 filed March 14, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 816 (now expired), and 60/372,381 filed April 12, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 818 (now expired); and

m) PCT Application Serial No. PCT/US03/30720 filed September 30, 2003 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 819CIP/PCT which claims the benefit of priority to U.S. Provisional Application Serial No. 60/416,186 filed October 02, 2002 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 819 (now expired); all of which are incorporated herein by reference in their entirety, specifically including, but not limited to, incorporation by reference of the tables in each application displaying sequence information, eMATRIX signatures, pfam signatures, signal peptide information, transmembrane domain information, chromosomal localization and tissue distribution information, 3-dimensional structural information and ancillary information. The material submitted on the compact discs contain the files labeled "824CIP PCT Table 9A.txt" – 128 kB (131,072 bytes), "824CIP PCT Table 9B.txt" – 440 kB (450,560 bytes), and saved on an IBM-PC, Windows 2000 operating system on March 17, 2004 at 8:28:45 PM and 9:51:26 PM, respectively and are all incorporated herein by reference in their entirety.

1.2 SEQUENCE LISTING

The sequences of the polynucleotides and polypeptides of the invention are listed in the Sequence Listing and are submitted on a compact disc containing the file labeled "824CIP PCT.txt"— 4.43 MB (4,653,056 bytes) which was created on an IBM PC, Windows 2000 operating system on March 23, 2004 at 10:29:33 AM. The Sequence Listing entitled "824CIP PCT.txt" is herein incorporated by reference in its entirety. A computer readable format ("CRF") and three duplicate copies ("Copy 1," "Copy 2" and "Copy 3") of the Sequence Listing "824CIP PCT.txt" are submitted herein. Applicants hereby state that the content of the CRF and Copies 1, 2, and 3 of the Sequence Listing, submitted in accordance with 37 CFR §1.821(c) and (e), respectively, are the same.

2. BACKGROUND OF THE INVENTION

2.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, circulating soluble factors, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-235, or 471-810 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases or unknown. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-235, or 471-810 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-235, or 471-810. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-235, or 471-810 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-235, or 471-810. The sequence information can be a segment of any one of SEQ ID NO: 1-235 or 471-810 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-235, or 471-810.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences;

and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-235, or 471-810 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-235, or 471-810 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-235, or 471-810; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-235, or 471-810; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-235, or 471-810. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-235, or 471-810; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in SEQ ID NO: 1-235, or 471-810; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homologue (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in SEQ ID NO: 236-470, or 811-1150, or Tables 3A, 3B, 4, 6, 9A, or 9B.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention (SEQ ID NO: 236-470, or 811-1150) also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-235, or 471-

810; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such processes is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or

quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The

invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can affect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention (SEQ ID NO: 236-470, or 811-1150) and the polynucleotides encoding them (SEQ ID NO: 1-235, or 471-810) are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Tables 2A and 2B); for which they have a signature region (as set forth in Tables 9A and 9B); or for which they have homology to a gene family (as set forth in Tables 3A and 3B). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only certain portion(s) of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the

sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G, or T (U) or unknown. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NO: 1-235, or 471-810.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-235, or 471-810. The sequence information can be a segment of any one of SEQ ID NO: 1-235, or 471-810 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-235, or 471-810, or those segments identified in Tables 3A, 3B, 4, 6, 9A, or 9B. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full-length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and

minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may

not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of

determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a new stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-235, or 471-810; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 1-235, or 471-810; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-235, or 471-810. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID

NO: 1-235, or 471-810; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing, or Tables 3A, 3B, 4, 6, 9A, or 9B; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homologue of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 236-470, or 811-1150 (for example, as set forth in Tables 3A, 3B, 4, 6, 9A, or 9B). Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-235, or 471-810 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-235, or 471-810 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-235, or 471-810 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-235, or 471-810, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-235, or 471-810, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-235, or 471-810 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology results for the nucleic acids of the present invention, including SEQ ID NO: 1-235, or 471-810 can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST (Basic Local Alignment Search Tool) program is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using FASTXY algorithm may be performed.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of

the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention could be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-235, or 471-810, or functional equivalents thereof, may be used to generate recombinant DNA molecules that

direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-235, or 471-810 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-235, or 471-810 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example: Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman,

Methods in Enzymology 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from

commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., Nat. Biotech 17, 870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-235, or 471-810, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1-235, or 471-810 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-235, or 471-810 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding

region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-235, or 471-810, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil,

(acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of

cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1-235, or 471-810). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair

mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and

B. subtilis. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial

strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, and regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.,* inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property

of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 236-470, or 811-1150 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-235, or 471-810 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NO: 1-235, or 471-810 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 236-470, or 811-1150 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 236-470, or 811-1150 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 236-470, or 811-1150.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites. Fragments are also identified in Tables 3A, 3B, 4, 6, 9A, or 9B.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained and confirmed by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell and sequencing of the cleaved product. One of skill in the art will recognize that the actual cleavage site may be different than that predicted. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed (See, e.g., Sakal et al., Prep. Biochem. Biotechnol. (2000), 30(2), pp. 107-23, incorporated herein by reference).

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The

synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in

Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 236-470, or 811-1150.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.*, as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for

biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by

reference), Pfam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), the GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95, 13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) incorporated herein by reference). Polypeptide sequences were examined by a proprietary algorithm, SeqLoc that separates the proteins into three sets of locales: intracellular, membrane, or secreted. This prediction is based upon three characteristics of each polypeptide, including percentage of cysteine residues, Kyte-Doolittle scores for the first 20 amino acids of each protein, and Kyte-Doolittle scores to calculate the longest hydrophobic stretch of the said protein. Values of predicted proteins are compared against the values from a set of 592 proteins of known cellular localization from the Swissprot database (Boeckmann *et al., Nucl. Acids Res.* 31:365-370 (2003) herein incorporated by reference in its entirety). Predictions are based upon the maximum likelihood estimation.

Presence of transmembrane region can be detected using the TMpred program (Hofmann and Stoffel, *Biol. Chem. Hoppe-Seyler* 374:166 (1993) herein incorporated by reference in its entirety).

The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention

and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety

(*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of

the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.,* inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple

deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model

systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The

homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on

gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or

amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of

mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or

pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create

cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the

invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast

activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from

chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and

disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of

an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro

antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of

localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis

Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention

(including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987)

Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening

assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.,* increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The responses of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an

inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include

but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival

or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution,

change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that

hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active

ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti- inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA,

latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water,

petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch,

potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active

compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also

may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as

micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into

the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question.

These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any

compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 236-470, or 811-1150, or Tables 3A, 3B, 4, 6, 9A, or 9B, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine

binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known

in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256, 495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107, 220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as

a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.13.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321, 522-525 (1986); Riechmann et al., Nature, 332, 323-327 (1988); Verhoeyen et al., Science, 239, 1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion

of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2, 593-596 (1992)).

4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80, 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227, 381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368, 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13, 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains

in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246, 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305, 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10, 3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121, 210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229, 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175, 217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical

coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5), 1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90, 6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152, 5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147, 60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176, 1191-1195 (1992) and Shopes, J. Immunol., 148, 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53, 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3, 219-230 (1989).

4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture

comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-235, or 471-810 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-235, or 471-810 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and

data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,

but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix-see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 15241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-235, or 471-810, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"

when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Okano, J. Neurochem. 56, 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-235, or 471-810. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NO: 1-235, or 471-810 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well-known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal

map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

4.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6), 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8), 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridgeheads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as

immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/μl) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-$MeIm_7$), is then added to a final concentration of 10 mM 1-$MeIm_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 μl/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-$MeIm_7$, is made fresh and 25 μl added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995), 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res., 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1), 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Nat'l. Acad. Sci., USA 91(11), 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24), 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids

Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *CviJI* normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*CviJI***), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *CviJI*** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *CviJI*** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 μg instead of 2-5 μg); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In

one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5 EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon

membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences.

5.2 EXAMPLE 2

Assemblage of Novel Nucleic Acids

The contigs or nucleic acids of the present invention, designated as SEQ ID NO: 473-815 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene, and exons from public domain genomic sequences predicated by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above databases that would extend the assemblage. Further, inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

5.3 EXAMPLE 3

Novel Nucleic Acids

The novel nucleic acids of the present invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequences were checked using FASTY and/or BLAST against Genebank (i.e., dbEST, gb pri, UniGene, and Genpept) and the Geneseq (Derwent). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NO: 1-470.

Table 1 shows the various tissue sources of SEQ ID NO: 1-236.

The homologs for polypeptides SEQ ID NO: 236-470, that correspond to nucleotide sequences SEQ ID NO: 1-235 were obtained by a BLASTP version 2.0al 19MP-WashU searches against Genpept and Geneseq (Derwent) using BLASTP algorithm. The results showing homologues for SEQ ID NO: 236-470 from Genpept 129 are shown in Tables 2A and 2B.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6, 219-235 (1999), herein incorporated by reference), all the polypeptide sequences were examined to determine whether they had identifiable signature regions. Scoring matrices of the eMatrix software package are derived from the BLOCKS, PRINTS, PFAM, PRODOM, and DOMO databases. Tables 9A and B herein submitted on compact disc as "824CIP PCT Table 9A.txt" and "824CIP PCT Table 9B.txt" and incorporated by reference in their entirety, show the accession number of the homologous eMatrix signature found in the indicated polypeptide sequence, its description, and the results obtained which include accession number subtype; raw score; p-value; and the position of signature in amino acid sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Tables 3A and B shows the name of the Pfam model found, the description, the p-value, and the Pfam score for the identified model within the sequence using Pfam version 7.2.

Table 4 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical

University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al. reference, was obtained for the polypeptide sequences.

Table 5 correlates nucleotide sequences of the invention to a specific chromosomal location when assignable.

Table 6 shows the number of transmembrane regions, their location(s), and TMPred score obtained, for each of the SEQ ID NO: 236-470 that had a TMPred score of 500 or greater, using the TMpred program (Hofman and Stoffel, Biol. Chem. Hoppe-Seyler 374:166 (1993), incorporated herein by reference).

Table 7 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-235, their corresponding polypeptide sequences SEQ ID NO: 236-470, their corresponding priority contig nucleotide sequences SEQ ID NO: 471-810, their corresponding priority contig polypeptide sequences SEQ ID NO: 811-1150, and the US serial number of the priority application (all of which are herein incorporated in their entirety), in which the contig sequence was filed.

Table 8 is a correlation table of the polynucleotide and polypeptide sequences SEQ ID NO: 1-1150 and their corresponding SEQ ID NO: in the priority U.S. Provisional Application, 60/458,824, from which the instant application claims the benefit of priority.

TABLE 1

Tissue Origin	Library/RNA Source	Nuvelo Library Name	SEQ ID NOS:
adult brain		AB2002	94 95
adult brain	GIBCO	AB3001	17 45 46 53 60 66 81 87 89 91 94 95 99 102 105 116 169 171 194
adult brain	GIBCO	ABD003	1 2 10 17 20 21 22 24 25 34 45 46 47 51 52 53 56 57 60 66 80 81 89 90 91 99 100 105 107 116 120 121 123 130 131 138 140 146 165 166 171 179 192 194 196
adult brain	Clontech	ABR001	13 16 25 34 56 66 79 100 116 117 128 169 189 193 194 199
adult brain	Clontech	ABR006	1 2 4 9 10 11 13 14 17 20 25 34 39 44 48 52 66 77 79 90 94 95 96 104 105 109 116 122 123 126 127 132 138 146 149 152 159 162 164 167 168 171 175 186 189 190 192 194 199 210 211
adult brain	Clontech	ABR008	1 2 8 9 11 13 14 17 20 40 41 42 43 44 45 46 47 48 52 60 66 69 71 77 79 82 83 85 92 93 94 95 100 101 102 103 104 107 109 110 117 123 124 128 133 138 140 141 142 143 146 147 149 159 161 162 164 168 169 171 173 174 179 190 193 196 199 204 205 206 210 211
adult brain	Clontech	ABR011	38 89
adult brain	BioChain	ABR013	35 36 66
adult brain	Invitrogen	ABR014	34 44 66 94 95 102
adult brain	Invitrogen	ABR015	66 89 142 143
adult brain	Invitrogen	ABR016	34 35 66 119 123 192
adult brain	Invitrogen	ABT004	8 25 34 37 51 56 60 71 81 83 101 103 128 141 147 149 161 192
cultured preadipocytes	Stratagene	ADP001	2 11 13 16 24 37 48 60 72 87 92 94 95 100 122 152 159 169 192 204 205 206 213 214 218
adrenal gland	Clontech	ADR002	1 11 13 16 17 20 21 24 36 38 50 53 57 60 69 81 82 84 87 89 93 94 95 102 105 117 124 132 137 138 146 147 159 168 191 194 195 205 206 210
adult heart	GIBCO	AHR001	5 11 17 21 24 25 45 46 48 50 52 53 57 58 66 68 72 78 80 81 82 85 86 89 93 94 95 100 101 103 116 117 125 130 131 146 160 161 168 169 171 176 179 193 194 195 199
adult kidney	GIBCO	AKD001	1 4 5 11 17 21 22 23 24 25 33 34 37 47 50 52 53 57 66 71 79 81 86 87 89 91 93 94 95 100 102 103 104 105 116 118 120 121 128 137 138 146 147 148 152 159 160 167 168 169 171 179 188 192 194 195 200
adult kidney	Invitrogen	AKT002	1 5 8 13 16 17 22 25 26 35 36 37 48 49 50 56 71 80 89 90 93 94 95 103 104 116 130 131 141 146 147 161 167 168 185 188 195 196 199 212
adult lung	GIBCO	ALG001	5 13 16 17 20 22 24 50 53 57 72 79 80 81 89 91 94 95 100 110 130 131 138 141 146 151 173 189 191 193 195
lymph node	Clontech	ALN001	3 27 35 36 57 79 80 91 116 130 131 138 159
young liver	GIBCO	ALV001	1 5 24 36 45 46 48 52 53 66 71 81 91 93 94 95 96 100 102 104 105 117 124 152 167 192 196
adult liver	Invitrogen	ALV002	1 17 22 24 26 35 36 43 45 46 50 60 82 90 93 94 95 96 100 104 124 141 146 147 149 152 154 155 156 160 162 164 167 168 188 196 200 219
adult liver	Clontech	ALV003	2 17 82 96 105 107 124 146 152 154 155 156 167 168 169 219
adult ovary	Invitrogen	AOV001	1 2 5 7 8 10 13 16 17 18 19 20 21 24 25 26 27 34 35 36 37 39 45 46 47 49 50 51 52 53 57 60 66 68 69 71 79 80 81 83 85 87 89 91 93 94 95 99 100 102 103

TABLE 1

Tissue Origin	Library/RNA Source	Nuvelo Library Name	SEQ ID NOS:
			105 109 116 118 124 128 129 130 131 137 138 141 145 146 147 149 150 152 153 161 168 171 172 179 181 192 194 196 199 210 211 213 217 218
adult placenta	Clontech	APL001	6 13 14 43 53 66 81 149 184 201 202 203
placenta	Invitrogen	APL002	2 6 24 35 107 119 138 212
adult spleen	Clontech	SPLc01	3 17 20 25 28 29 30 31 32 35 36 52 66 69 79 81 85 87 89 94 95 100 103 105 128 146 168 169 170 191 222
adult spleen	GIBCO	ASP001	2 3 4 20 25 27 28 29 30 31 32 35 36 48 51 52 53 66 73 74 75 76 81 89 94 95 100 103 105 132 137 158 169 222
adult testis	GIBCO	ATS001	5 17 20 24 37 52 53 54 55 57 60 66 81 89 99 100 103 109 116 139 149 161 171 179 222
adult bladder	Invitrogen	BLD001	13 16 17 35 36 44 48 60 69 71 81 105 128 146 152 195 200 213
bone marrow	Clontech	BMD001	4 5 11 18 19 20 21 24 27 35 36 37 38 48 52 53 57 66 79 81 82 87 89 91 93 94 95 100 103 104 105 109 116 117 137 138 140 145 146 147 148 150 159 168 171 172 173 179 183 195 213
bone marrow	GF	BMD002	1 2 3 8 11 17 20 21 22 25 28 29 30 31 32 34 35 36 37 43 48 52 53 57 60 61 66 81 85 86 93 94 95 100 102 103 104 109 110 117 136 137 138 146 148 153 157 158 159 167 168 169 170 171 185 195 196 204 205 206
bone marrow	CD34+ cells	STM001	94 95 194
bone marrow	Clontech	BMD004	35 36 213
bone marrow	Clontech	BMD007	35 36
adult colon	Invitrogen	CLN001	17 22 24 45 46 60 71 89 103 105 110 146 159 160 169 195 196
mix	B/I/C	CTL016	167 213
mixed		CTL021	36 188 195
adult cervix	BioChain	CVX001	1 5 9 11 13 16 17 20 21 22 24 27 34 39 40 41 42 50 51 53 56 57 60 66 79 81 84 89 91 93 94 95 100 103 105 116 118 119 122 128 133 137 138 140 141 145 146 147 149 153 159 161 168 171 173 179 191 193 194 207 208 209 218 223
lymphocyte	CA46 cells	DGD001	71 94 95 105 195 210
diaphragm	BioChain	DIA002	103 146
endothelial cells	Stratagene	EDT001	5 7 8 9 11 17 20 21 22 24 25 34 37 45 46 48 50 51 52 53 60 66 79 81 82 84 85 86 87 89 93 94 95 102 104 128 138 140 141 146 147 149 161 168 169 171 176 179 192 195 217
esophagus	BioChain	ESO002	93
fetal brain	Clontech	FBR001	17 25 34 56 66 70 152 189 196
fetal brain	Clontech	FBR004	52 57 77 79 98
fetal brain	Clontech	FBR006	1 8 9 11 13 14 17 20 21 26 40 41 42 43 47 48 52 53 60 66 69 71 79 81 82 83 85 87 89 93 94 95 103 105 107 108 110 117 121 122 123 138 140 141 146 153 159 161 168 171 177 178 190 196 198 204 205 206 211
fetal brain	Clontech	FBRs03	94 95
fetal brain	Invitrogen	FBT002	17 24 26 34 38 56 60 84 94 95 100 117 120 121 126 127 147 149 152 168 169
fetal heart	Invitrogen	FHR001	8 17 48 52 57 60 66 71 79 85 87 89 93 94 95 100 101 103 104 118 119 136 137 150 160 161 168 171 179 181 186 195 204 205 206 210 213

TABLE 1

Tissue Origin	Library/RNA Source	Nuvelo Library Name	SEQ ID NOS:
fetal kidney	Clontech	FKD001	1 7 11 21 79 103 171 213
fetal kidney	Clontech	FKD002	2 8 20 45 46 48 60 61 66 79 85 94 95 100 104 110 118 139 171 180 181 188 191 196 204 205 206
fetal lung	Clontech	FLG001	1 5 20 24 45 46 57 69 71 85 116 150 160 199 213
fetal lung	Invitrogen	FLG003	13 16 20 24 45 46 56 60 71 79 84 85 94 95 109 128 146 149 161
fetal lung	Clontech	FLG004	94 95
fetal liver-spleen	Columbia University	FLS001	1 2 5 6 8 9 11 13 14 15 16 17 18 19 20 21 22 24 25 26 35 36 45 46 48 50 52 56 57 60 66 70 71 79 80 82 85 86 87 89 92 93 94 95 96 97 100 102 103 104 105 107 116 117 118 124 130 131 136 137 138 140 141 146 147 149 150 151 152 154 155 156 159 160 161 162 163 164 167 168 169 171 173 179 181 184 192 196 199 200 201 202 203 210 212 213 218 222 223 227
fetal liver-spleen	Columbia University	FLS002	1 2 5 6 7 8 9 11 13 14 16 17 18 19 20 21 24 25 28 29 30 31 32 34 35 36 37 43 45 46 48 51 52 56 57 59 60 69 70 71 81 85 86 87 89 93 96 97 98 100 102 103 104 105 107 116 117 124 136 137 138 146 147 149 150 152 153 154 155 156 159 161 162 164 165 166 167 169 171 173 176 179 182 184 191 192 193 194 195 196 199 200 201 202 203 210 212 213 217 219
fetal liver-spleen	Columbia University	FLS003	2 6 8 11 13 14 16 17 18 19 22 24 48 60 71 80 86 87 96 99 100 102 104 117 130 131 137 141 154 155 156 167 184 187 194 195 201 202 203 218 222
fetal liver	Invitrogen	FLV001	24 26 45 46 71 80 82 124 130 131 136 149 160 167 168 173 182 195 200 212
fetal liver	Clontech	FLV002	17 26 43 44 45 46 77 96 98 117 137 152 167 182 200
fetal liver	Clontech	FLV004	8 11 21 25 27 37 45 46 48 71 79 85 86 89 93 94 95 96 104 107 124 154 155 156 162 164 167 183 213
fetal muscle	Invitrogen	FMS001	11 26 48 52 67 125 128 140 141 149 160 169 191 213
fetal muscle	Invitrogen	FMS002	2 11 17 20 22 24 37 48 52 53 57 60 66 81 89 94 95 100 103 110 117 125 140 141 146 149 153 159 160 168 171 194 200 213
fetal skin	Invitrogen	FSK001	1 7 13 16 20 24 25 26 34 40 41 42 43 50 57 60 66 71 79 84 87 89 93 94 95 100 101 102 103 107 117 119 122 125 126 127 128 140 147 148 149 159 169 189 191 193 205 206 208 209 213
fetal skin	Invitrogen	FSK002	13 14 17 20 21 48 50 61 63 64 65 71 78 84 87 93 94 95 100 102 105 116 122 126 127 132 137 140 149 159 161 168 171 191 204 205 206 213 214 218
umbilical cord	BioChain	FUC001	1 5 7 13 16 17 20 21 37 38 43 53 60 71 78 80 81 89 103 122 128 130 131 146 147 149 150 168 171 173 187 193 199 210 212 213 217 218
fetal brain	GIBCO	HFB001	1 4 5 10 11 12 17 22 25 37 38 39 52 53 58 59 60 66 81 84 85 87 89 90 91 105 116 118 122 135 145 146 150 152 159 162 164 169 171 179 181 182 189 192 194 196 199 211
macrophage	Invitrogen	HMP001	21 22 51 82 89 94 95 100 148 167 169
infant brain	Columbia University	IB2002	1 2 4 11 17 21 25 26 38 40 41 42 44 47 48 52 56 58 60 61 66 71 77 79 82 84 87 91 93 94 95 102 103 104 107 108 113 116 117 120 121 122 123 135 138 159 161 162 164 168 173 174 188 189 192 199 200 211 217

TABLE 1

Tissue Origin	Library/RNA Source	Nuvelo Library Name	SEQ ID NOS:
infant brain	Columbia University	IB2003	10 11 17 25 26 34 38 40 41 42 44 47 48 56 60 61 66 71 79 80 81 84 87 93 94 95 102 103 105 107 113 116 117 120 121 122 128 130 131 135 138 141 146 160 161 168 176 188 199 200 213 218
infant brain	Columbia University	IBM002	44 47 52 58 60 135 159
infant brain	Columbia University	IBS001	11 48 66 77 84 91 94 117
lung, fibroblast	Stratagene	LFB001	13 16 20 37 66 81 83 89 91 105 116 128 147 161 168 173 179 218
lung tumor	Invitrogen	LGT002	1 2 4 7 13 16 17 20 21 24 35 36 37 43 44 52 57 59 60 66 71 80 82 87 88 89 91 93 94 95 97 99 100 105 106 107 130 131 137 138 139 141 142 143 144 145 146 147 149 150 153 162 164 167 168 171 179 194 195 199 212 213 215 216 218 227
lymphocytes	ATCC	LPC001	2 11 17 20 22 24 25 27 43 48 52 57 66 71 80 85 87 89 93 102 103 109 111 112 117 130 131 139 157 158 161 168 172 194 195 215 216
leukocyte	GIBCO	LUC001	3 8 9 11 17 18 19 20 21 22 24 25 27 28 29 30 31 32 35 36 37 45 46 48 52 53 57 60 61 66 71 73 74 75 76 80 81 82 85 89 91 93 94 95 97 100 102 103 104 105 107 117 128 130 131 136 137 138 139 145 146 147 150 157 158 159 161 167 168 169 171 172 179 183 191 194 195 199 212 217 218
leukocyte	Clontech	LUC003	27 43 52 85 146 222
melanoma from-cell-line-ATCC-#CRL-1424	Clontech	MEL004	7 17 21 60 71 89 138 141 159 179
mammary gland	Invitrogen	MMG001	1 2 5 8 13 16 17 21 22 24 25 26 28 29 30 31 32 34 35 36 45 46 47 53 60 61 62 63 64 65 66 71 79 80 81 82 83 84 89 90 93 94 95 97 100 103 107 122 128 130 131 138 139 141 144 147 149 150 152 153 159 160 169 172 176 179 191 192 195 199 213 218
induced neuron-cells	Stratagene	NTD001	43 52 56 103 107 168
retinoic acid-induced-neuronal-cells	Stratagene	NTR001	43 60 122 146 171
neuronal cells	Stratagene	NTU001	11 13 16 60 61 82 122 176 179 200
mixed		CGSP006	68 92
Mixed		CGSd001	35 36 78 141 187
Mixed		CGSd002	222
Mixed		CGSd003	50 121
Mixed		CGSd004	6 50 111 112
Mixed		CGSd005	50 135 168 169
Mixed		CGSd006	3 18 19 23 35 36 44 50 51 60 69 80 82 87 94 95 108 118 130 131 135 136 159 160 161 165 166 168 169 185 188 200
Mixed		CGSd009	3 20 23 35 36 44 52 57 69 78 80 82 84 89 94 95 116 118 120 121 130 131 136 160 165 166 168 204 205 206 213
Mixed		CGd007	35 36 50 80 124 130 131 136 160 168 181 211 222
Mixed		CGd008	1 26 35 36 50 80 100 124 130 131 168 181 211 222
mixed	EST clones	CGd010	26 35 36 44 50 87 127 129 160 165 166 168
mixed		CGd011	11 35 36 38 48 52 94 95 104 117 163 164

TABLE 1

Tissue Origin	Library/RNA Source	Nuvelo Library Name	SEQ ID NOS:
mixed		CGd012	2 3 11 13 14 15 16 23 24 26 35 36 38 40 41 42 43 48 49 51 52 53 57 58 59 60 62 63 64 65 72 79 81 82 87 93 94 95 100 101 102 103 104 105 107 115 117 118 119 122 123 124 125 126 127 139 146 147 148 149 161 162 163 164 167 168 169 173 190 194 195 196 201 202 203 205 206 217 218
mixed		CGd013	24 35 48 49 59 62 63 64 65 88 107 118 119 122 147 149 168 169 195 215 216
mixed		CGd015	2 4 5 6 17 24 35 36 39 45 46 53 56 66 81 82 89 91 94 95 96 99 138 146 153 159 162 164 167 168 169 171 199 201 202 213
mixed		CGd016	2 4 13 14 17 35 36 40 41 42 50 53 56 66 81 86 105 122 138 142 143 146 147 149 153 159 161 163 164 168 175 193 219 227
mixed		CGd021	2 11 13 16 35 36 61 72 77 80 90 100 126 127 130 131 160 162 164 213
mixed		CGd022	94 95 110
mixed	PCR products	PCR2V1	13 14 15 16 53 66 67 68 81 82 92 116 122 146 213 222
Mix	B/I/C	SUP002	1 2 8 22 35 43 56 60 66 79 80 81 83 94 95 99 100 109 116 130 131 139 161 167 175 200 210 213
mix	B/I/C	SUP005	35 36 60 213
mix	B/I/C	SUP008	50 66 94 95 167 213
mix	B/I/C	SUP009	35 52 94 95 96 167
mixed		PGEMV1	13 15 16 36 45 46 52 63 64 65 66 68 69 87 89 92 94 95 116 122 125 137 192 213 222
pituitary gland	Clontech	PIT004	6 11 25 38 92 100 103 105 147 168 179 199 205 206
placenta	Clontech	PLA003	1 6 13 14 15 16 17 21 24 48 66 71 79 81 85 87 89 94 95 100 119 133 137 146 162 164 168 171 184 186 201 202 203 210 212
prostate	Clontech	PRT001	4 11 17 24 36 53 55 57 66 81 89 90 94 95 96 100 102 125 138 161 182 194 195 199
rectum	Invitrogen	REC001	1 25 34 35 36 66 71 84 94 95 105 147 180 188
salivary gland	Clontech	SAL001	17 24 35 53 57 81 86 89 94 95 105 138 141 194
small intestine	Clontech	SIN001	1 2 13 16 17 20 21 23 24 25 27 35 37 43 47 50 52 53 57 60 61 72 73 74 75 76 77 80 81 82 87 89 93 94 95 102 103 109 111 112 117 130 131 138 146 147 153 159 168 173 176 191 192 196
skeletal muscle	Clontech	SKM001	25 61 89 93 94 95 103 117 147 160 192 212
spinal cord	Clontech	SPC001	17 20 24 27 47 48 52 53 57 60 66 81 87 89 90 94 95 105 107 117 138 146 149 150 159 161 162 164 168 171 193 194
stomach	Clontech	STO001	1 4 5 17 20 23 27 35 53 81 89 94 95 100 103 105 159 168 195
thalamus	Clontech	THA002	34 38 44 45 46 94 95 100 101 102 135 146 160 171 199
thymus	Clontech	THM001	8 13 16 21 24 35 36 45 46 57 71 87 89 91 94 95 103 105 117 133 138 139 149 150 153 159 168 173 195 222
thymus	Clontech	THMc02	1 8 11 17 27 28 29 30 31 32 35 36 37 60 66 69 71 79 80 87 89 92 100 104 105 107 122 128 130 131 137 139 141 146 161 168 172 194 213 219
thyroid gland	Clontech	THR001	1 9 11 13 16 17 21 24 25 34 36 43 47 48 50 53 57 60 61 63 64 65 66 71 80 81 82 84 89 94 95 97 99 100

TABLE 1

Tissue Origin	Library/RNA Source	Nuvelo Library Name	SEQ ID NOS:
			101 102 103 104 105 109 116 117 124 128 130 131 132 137 138 140 146 153 160 161 168 169 171 176 194 196 199 217 218 219
trachea	Clontech	TRC001	1 8 22 35 36 40 41 42 53 57 66 69 71 81 82 105 107 115 116 128 138 159 173 195 196
uterus	Clontech	UTR001	7 21 50 52 53 60 72 81 89 94 95 100 101 103 122 138 148 159 169 194 197 208 209

*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human so\spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
236	gi18676574	2747	0.0	99	(AK074113) FLJ00184 protein [Homo sapiens]
236	gi20198487	5657	0.0	99	AF441771_1 (AF441771) 182kDa tankyrase1-binding protein [Homo sapiens]
236	gi28278261	941	1e-99	98	(BC046216) Similar to tankyrase 1 binding protein 1, 182kDa [Homo sapiens]
237	gi10998129	437	3e-41	37	(AP002040) ubiquitin carboxyl-terminal hydrolase-like protein [Arabidopsis thaliana]
237	gi27754270	437	3e-41	37	(BT002760) putative ubiquitin carboxyl-terminal hydrolase [Arabidopsis thaliana]
237	gi6671947	437	3e-41	37	AC016795_20 (AC016795) putative ubiquitin carboxyl-terminal hydrolase [Arabidopsis thaliana]
238	gi16506257	1652	0.0	99	AF329488_1 (AF329488) IFGP1 [Homo sapiens]
238	gi18140081	1640	0.0	99	AF459634_1 (AF459634) immunoglobulin superfamily receptor translocation associated 5 [Homo sapiens]
238	gi21707303	1640	0.0	99	(BC033690) Fc receptor-like protein 1 [Homo sapiens]
239	gi1372963	178	3e-12	68	(M85148) cytochrome oxidase subunit III [Macaca mulatta]
239	gi21104492	743	8e-78	100	(AB064665) OK/SW-CL.16 [Homo sapiens]
240	gi18088315	528	4e-53	100	AAH20623 (BC020623) chromosome 8 open reading frame 4 [Homo sapiens]
240	gi18203818	528	4e-53	100	AAH21672 (BC021672) chromosome 8 open reading frame 4 [Homo sapiens]
240	gi8745547	528	4e-53	100	AF268037_1 (AF268037) C8ORF4 protein [Homo sapiens]
241	gi12803759	1128	e-122	100	AAH02717 (BC002717) Similar to chorionic somatomammotropin hormone 1 (placental lactogen) [Homo sapiens]
241	gi13543526	1128	e-122	100	AAH05921 (BC005921) chorionic somatomammotropin hormone 1 (placental lactogen) [Homo sapiens]
241	gi18088830	1128	e-122	100	AAH20756 (BC020756) chorionic somatomammotropin hormone 1 (placental lactogen) [Homo sapiens]
242	gi13872813	3662	0.0	96	(AJ306906) fibulin-6 [Homo sapiens]
242	gi14575679	3662	0.0	96	AF156100_1 (AF156100) hemicentin [Homo sapiens]
242	gi3372528	608	3e-61	33	(AF051403) fibulin-1 isoform D precursor [Caenorhabditis elegans]
243	gi20149223	1097	e-118	100	AF493783_1 (AF493783) koyt binding protein 1 [Homo sapiens]
243	gi20149229	1097	e-118	100	AF493786_1 (AF493786) koyt binding protein 1 [Homo sapiens]
243	gi21105773	1094	e-118	99	AF512007_1 (AF512007) proline rich protein BCA3 [Homo sapiens]
244	gi15929192	1487	e-163	99	AAH15047 (BC015047) Unknown (protein for MGC:9522) [Homo sapiens]
244	gi16553200	1571	e-173	100	(AK057477) unnamed protein product [Homo sapiens]
244	gi23271139	1265	e-138	81	(BC035953) Similar to hypothetical protein FLJ32915 [Mus musculus]
245	gi8118086	6532	0.0	80	AF218940_1 (AF218940) formin-2 [Mus

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					musculus]
245	gi8118088	1715	0.0	100	(AF218941) formin 2-like protein [Homo sapiens]
245	gi8118090	1533	e-168	100	(AF218942) formin 2-like protein [Homo sapiens]
246	gi12584845	1783	0.0	99	AF284753_1 (AF284753) X2HRIP110 [Homo sapiens]
246	gi21619703	1643	0.0	99	(BC032561) Similar to retinoid x receptor interacting protein [Homo sapiens]
246	gi6523831	1800	0.0	100	AF113538_1 (AF113538) retinoid x receptor interacting protein [Homo sapiens]
248	gi11177164	16715	0.0	81	AF206329_1 (AF206329) polydom protein [Mus musculus]
248	gi12060830	2520	0.0	94	AF308289_1 (AF308289) serologically defined breast cancer antigen NY-BR-38 [Homo sapiens]
248	gi14198157	3176	0.0	79	(BC008135) polydomain protein [Mus musculus]
249	gi11177164	4047	0.0	83	AF206329_1 (AF206329) polydom protein [Mus musculus]
249	gi22536178	329	8e-29	27	(AF540378) SELE: selectin E (endothelial adhesion molecule 1) [Homo sapiens]
249	gi3115964	329	8e-29	27	(AL021940) dJ117P20.2 (E-Selectin precursor (CD62E, ELAM-1 Endothelial Leukocyte Adhesion Molecule 1, LECAM-2 Leukocyte-Endothelial Cell Adhesion Molecule 2)) [Homo sapiens]
250	gi11177164	1975	0.0	80	AF206329_1 (AF206329) polydom protein [Mus musculus]
250	gi499688	368	1e-33	57	(L33862) fibropellin III [Heliocidaris erythrogramma]
250	gi7297206	513	2e-50	33	(AE003615) CG9138-PA [Drosophila melanogaster]
251	gi11177164	12675	0.0	80	AF206329_1 (AF206329) polydom protein [Mus musculus]
251	gi12060830	2520	0.0	94	AF308289_1 (AF308289) serologically defined breast cancer antigen NY-BR-38 [Homo sapiens]
251	gi14198157	3176	0.0	79	(BC008135) polydomain protein [Mus musculus]
252	gi11037740	2130	0.0	97	(AF304118) apoptotic cell clearance receptor PtdSerR [Mus musculus]
252	gi22086529	1881	0.0	85	(AF401484) phosphatidylserine receptor long form [Danio rerio]
252	gi23491564	1950	0.0	89	(AB073711) phosphatidylserine receptor beta [Homo sapiens]
254	gi21615526	2413	0.0	98	(AJ314648) ATP(GTP)-binding protein [Homo sapiens]
255	gi15987495	2800	0.0	100	AF378757_1 (AF378757) tumor endothelial marker 7-related precursor [Homo sapiens]
255	gi15987503	2538	0.0	91	AF378761_1 (AF378761) tumor endothelial marker 7-related precursor [Mus musculus]
255	gi5457119	1287	e-140	99	AF154005_1 (AF154005) junction adhesion molecule [Homo sapiens]
256	gi12805505	958	e-102	97	(BC002229) Similar to CHMP1.5 protein [Mus musculus]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
256	gi17933108	972	e-104	100	AF306520_1 (AF306520) C18orf2 [Homo sapiens]
256	gi9885435	957	e-102	100	AF281064_1 (AF281064) CHMP1.5 [Homo sapiens]
257	gi17862416	820	3e-86	54	(AY069540) LD26422p [Drosophila melanogaster]
257	gi27353006	672	5e-69	41	(AP005952) bll4742 [Bradyrhizobium japonicum]
257	gi7291920	826	7e-87	50	(AE003467) CG7049-PA [Drosophila melanogaster]
258	gi13529161	908	8e-97	100	AAH05350 (BC005350) Similar to regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein) [Homo sapiens]
258	gi190979	908	8e-97	100	(M18963) islet regenerating protein [Homo sapiens]
258	gi5764555	908	8e-97	100	AF172331_1 (AF172331) lithostathine [Homo sapiens]
259	gi16551383	629	2e-64	100	AF403478_1 (AF403478) SIPL [Homo sapiens]
259	gi18087553	621	2e-63	62	AF462818_1 (AF462818) AT4g14710/dl3395c [Arabidopsis thaliana]
259	gi21555216	621	2e-63	62	(AY086754) submergence induced protein 2A [Arabidopsis thaliana]
260	gi19880264	1649	0.0	92	(AF363483) metallo phosphoesterase [Homo sapiens]
260	gi19880265	1649	0.0	92	(AF363483) metallo phosphoesterase [Homo sapiens]
260	gi19880267	1649	0.0	92	AF363484_1 (AF363484) metallo phosphoesterase [Homo sapiens]
261	gi15963593	7806	0.0	100	AF414401_1 (AF414401) ADAMTS13 [Homo sapiens]
261	gi16117338	7806	0.0	100	(AB069698) von Willebrand factor-cleaving protease [Homo sapiens]
261	gi16306598	7802	0.0	99	(AY055376) von Willebrand factor-cleaving protease precursor [Homo sapiens]
262	gi13021810	1349	e-147	100	AF291815_1 (AF291815) NK cell receptor [Homo sapiens]
262	gi20380757	1565	e-172	100	(BC027867) 19A24 protein [Homo sapiens]
262	gi7161175	1410	e-154	100	(AJ271869) 19A24 protein [Homo sapiens]
263	gi10141011	1798	0.0	55	(AF246701) leukocyte cell-surface molecule [Mus musculus]
263	gi10197717	3426	0.0	99	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
263	gi1235698	3180	0.0	97	(L42621) Ly-9 gene product [Homo sapiens]
264	gi10197717	216	3e-17	100	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
264	gi9588414	216	3e-17	100	(AL121985) bA404F10.5 (lymphocyte antigen 9) [Homo sapiens]
265	gi10141011	1735	0.0	54	(AF246701) leukocyte cell-surface molecule [Mus musculus]
265	gi10197717	3340	0.0	97	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
265	gi1235698	3216	0.0	99	(L42621) Ly-9 gene product [Homo sapiens]
266	gi10141011	1690	0.0	53	(AF246701) leukocyte cell-surface molecule [Mus musculus]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
266	gi10197717	3274	0.0	96	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
266	gi1235698	3028	0.0	93	(L42621) Ly-9 gene product [Homo sapiens]
267	gi10141011	1706	0.0	55	(AF246701) leukocyte cell-surface molecule [Mus musculus]
267	gi10197717	3216	0.0	99	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
267	gi1235698	3135	0.0	97	(L42621) Ly-9 gene product [Homo sapiens]
268	gi22003417	182	9e-13	39	AF394058_1 (AF394058) neogenin [Danio rerio]
268	gi27469556	246	3e-20	42	(BC042054) Similar to putative neuronal cell adhesion molecule [Homo sapiens]
268	gi3068592	234	9e-19	42	(AF026465) punc [Mus musculus]
269	gi13278924	748	3e-78	98	AAH04217 (BC004217) neural proliferation, differentiation and control, 1 [Homo sapiens]
269	gi18028281	748	3e-78	98	AF327349_1 (AF327349) NPDC-1 protein [Homo sapiens]
269	gi8515886	748	3e-78	98	AF272357_1 (AF272357) NPDC1-like protein [Homo sapiens]
270	gi14603095	1814	0.0	81	AAH10018 (BC010018) S-adenosylhomocysteine hydrolase [Homo sapiens]
270	gi15079562	1814	0.0	81	AAH11606 (BC011606) Similar to S-adenosylhomocysteine hydrolase [Homo sapiens]
270	gi15929766	1815	0.0	81	(BC015304) S-adenosylhomocysteine hydrolase [Mus musculus]
271	gi15559823	2253	0.0	89	AAH14258 (BC014258) Similar to immunoglobulin heavy constant gamma 3 (G3m marker) [Homo sapiens]
271	gi16741064	2135	0.0	85	AAH16381 (BC016381) Similar to immunoglobulin heavy constant gamma 3 (G3m marker) [Homo sapiens]
271	gi17939658	2145	0.0	86	AAH19337 (BC019337) Similar to immunoglobulin heavy constant gamma 3 (G3m marker) [Homo sapiens]
272	gi11493982	303	4e-27	70	AF208232_1 (AF208232) TLH29 protein precursor [Homo sapiens]
272	gi15929988	497	1e-49	100	AAH15423 (BC015423) Similar to TLH29 protein precursor [Homo sapiens]
272	gi21618549	303	4e-27	70	(BC032626) TLH29 protein precursor [Homo sapiens]
273	gi21961553	1998	0.0	98	(BC034781) neuronal pentraxin II [Homo sapiens]
273	gi881934	2013	0.0	98	(U26662) neuronal pentraxin II [Homo sapiens]
273	gi9931976	2013	0.0	98	(U29195) neuronal pentraxin II [Homo sapiens]
274	gi1333929	161	2e-10	39	(X66285) HC1 ORF [Mus musculus]
274	gi21928439	166	4e-11	32	(AB065580) seven transmembrane helix receptor [Homo sapiens]
274	gi862343	161	2e-10	36	(L10908) Gcap1 gene product [Mus musculus]
275	gi14280020	3380	0.0	49	(AF312825) collagen type XX alpha 1 precursor [Gallus gallus]
275	gi288873	1294	e-140	36	(X70793) collagen XIV [Gallus gallus]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
275	gi288875	1294	e-140	36	(X70792) collagen XIV [Gallus gallus]
276	gi14280020	3652	0.0	52	(AF312825) collagen type XX alpha 1 precursor [Gallus gallus]
276	gi288873	1294	e-140	36	(X70793) collagen XIV [Gallus gallus]
276	gi288875	1294	e-140	36	(X70792) collagen XIV [Gallus gallus]
277	gi14280020	3465	0.0	50	(AF312825) collagen type XX alpha 1 precursor [Gallus gallus]
277	gi288873	1294	e-140	36	(X70793) collagen XIV [Gallus gallus]
277	gi288875	1294	e-140	36	(X70792) collagen XIV [Gallus gallus]
278	gi12653223	876	2e-92	42	AAH00380 (BC000380) DNA segment on chromosome 21 (unique) 2056 expressed sequence [Homo sapiens]
278	gi2258274	876	2e-92	42	(U79775) NNP-1/Nop52 [Homo sapiens]
278	gi7768761	876	2e-92	42	(AP001752) NNP-1/Nop52 (NNP-1), novel nuclear protein 1 [Homo sapiens]
279	gi20975686	2911	0.0	100	(AJ487518) leucine-rich glioma inactivated protein 3 [Homo sapiens]
279	gi21359658	2911	0.0	100	(AF467956) LGI3 [Homo sapiens]
279	gi21901937	2911	0.0	100	(AJ487961) LGI1-like protein 4 [Homo sapiens]
281	gi15079633	226	3e-17	25	AAH11634 (BC011634) Similar to G protein-coupled receptor 30 [Homo sapiens]
281	gi1707500	226	3e-17	25	(Y08162) heptahelix receptor [Homo sapiens]
281	gi1894789	226	3e-17	25	(X98510) G protein-coupled receptor [Homo sapiens]
282	gi23271350	651	3e-66	41	(BC036360) Similar to chondroadherin [Homo sapiens]
282	gi470672	653	2e-66	41	(U08018) cartilage leucine-rich protein [Bos taurus]
282	gi6572272	4157	0.0	100	(AL035681) dJ756G23.1 (novel Leucine Rich Protein) [Homo sapiens]
283	gi22347831	1028	e-110	42	(AF533250) zinc finger protein [Homo sapiens]
283	gi27371193	968	e-103	44	(BC041661) zinc finger protein 305 [Homo sapiens]
283	gi36603	2198	0.0	99	(Z11773) SRE-ZBP [Homo sapiens]
284	gi19171150	1130	e-121	54	(AJ311903) ADAMTS18 protein [Homo sapiens]
284	gi19171178	3590	0.0	79	(AJ315734) metalloprotease disintegrin 16 with thrombospondin type I motif [Homo sapiens]
284	gi5923786	1140	e-123	34	AF140674_1 (AF140674) zinc metalloprotease ADAMTS6 [Homo sapiens]
285	gi21724166	1093	e-118	100	(AY039241) gastric cancer antigen Ga34 [Homo sapiens]
285	gi6252444	1282	e-140	99	(AB034695) endomucin-2 [Homo sapiens]
285	gi8547215	1289	e-141	100	AF205940_1 (AF205940) endomucin [Homo sapiens]
286	gi17862986	777	6e-81	44	(AY069825) SD07339p [Drosophila melanogaster]
286	gi21320872	2744	0.0	87	(AB041610) Cog8 [Mus musculus]
286	gi7297851	1143	e-123	43	(AE003632) CG6488-PA [Drosophila melanogaster]
287	gi18848244	3785	0.0	96	(BC024131) similar to metastasis suppressor protein [Mus musculus]
287	gi27769040	1848	0.0	94	(BC042632) Similar to cDNA sequence

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					BC024131 [Mus musculus]
287	gi6539606	3918	0.0	99	(AF086645) metastasis suppressor protein [Homo sapiens]
288	gi12406754	446	1e-43	100	(AX061647) unnamed protein product [Homo sapiens]
288	gi18378673	446	1e-43	100	AF462605_1 (AF462605) PATE [Homo sapiens]
289	gi12406754	607	4e-62	89	(AX061647) unnamed protein product [Homo sapiens]
289	gi18378673	608	3e-62	90	AF462605_1 (AF462605) PATE [Homo sapiens]
290	gi12406754	691	9e-72	99	(AX061647) unnamed protein product [Homo sapiens]
290	gi18378673	692	7e-72	100	AF462605_1 (AF462605) PATE [Homo sapiens]
291	gi23092843	209	1e-15	37	(AE003475) CG16757-PA [Drosophila melanogaster]
291	gi2623757	334	4e-30	42	(U72994) neurabin [Rattus norvegicus]
291	gi3598728	355	2e-32	44	(AC004022) Neurabin-like; similar to U72994 (PID:g2623757) [Homo sapiens]
292	gi27802717	2872	0.0	52	(AL627263) SI:bZ1L9.1 (novel protein similar to ATPase, Class I, type 8B, member 1 (ATP8B1)) [Danio rerio]
292	gi6457274	3340	0.0	56	AF156551_1 (AF156551) putative E1-E2 ATPase [Mus musculus]
292	gi7715417	5114	0.0	85	AF236061_1 (AF236061) RING-finger binding protein [Oryctolagus cuniculus]
293	gi18496661	2676	0.0	100	(AF465770) copine-like protein isoform A [Homo sapiens]
293	gi18496663	2676	0.0	100	(AF465771) copine-like protein isoform B [Homo sapiens]
293	gi23271332	1921	0.0	72	(BC035334) Similar to copine VII [Homo sapiens]
294	gi1915909	11411	0.0	95	(X99805) alpha tectorin [Mus musculus]
294	gi3309151	11773	0.0	99	(AF055136) alpha-tectorin [Homo sapiens]
294	gi4049439	8659	0.0	73	(AJ012287) alpha tectorin [Gallus gallus]
295	gi161467	1326	e-144	38	(L08692) fibropellin Ia [Strongylocentrotus purpuratus]
295	gi18676472	7210	0.0	99	(AK074062) FLJ00133 protein [Homo sapiens]
295	gi18676498	2724	0.0	89	(AK074075) FLJ00146 protein [Homo sapiens]
296	gi23172107	139	1e-07	36	(AE003745) CG5926-PA [Drosophila melanogaster]
297	gi24636593	204	1e-14	28	(AB095109) CiGl [Ciona intestinalis]
297	gi28279424	181	5e-12	55	(BC045743) Similar to g1-related zinc finger protein [Homo sapiens]
297	gi5441942	1723	0.0	100	AC004997_5 (AC004997) supported by mouse EST AA538043 (NID:g2284036) [Homo sapiens]
298	gi20086516	490	3e-48	100	AF245303_1 (AF245303) prominin-2 variant A [Homo sapiens]
298	gi20086518	490	3e-48	100	AF245304_1 (AF245304) prominin-2 variant B [Homo sapiens]
298	gi24637566	300	3e-26	50	(AF508942) prominin-2 [Rattus norvegicus]
299	gi20086516	3442	0.0	99	AF245303_1 (AF245303) prominin-2 variant

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					A [Homo sapiens]
299	gi20086518	3442	0.0	99	AF245304_1 (AF245304) prominin-2 variant B [Homo sapiens]
299	gi24637566	2646	0.0	75	(AF508942) prominin-2 [Rattus norvegicus]
300	gi20086516	1063	e-114	99	AF245303_1 (AF245303) prominin-2 variant A [Homo sapiens]
300	gi20086518	1063	e-114	99	AF245304_1 (AF245304) prominin-2 variant B [Homo sapiens]
300	gi24637566	787	1e-82	75	(AF508942) prominin-2 [Rattus norvegicus]
301	gi14714659	386	6e-37	100	AAH10469 (BC010469) Similar to homolog of mouse MAT-1 oncogene [Homo sapiens]
301	gi473910	141	2e-08	90	(L31958) mammary transforming protein [Mus musculus]
301	gi598187	310	4e-28	82	(L37385) unknown [Homo sapiens]
302	gi13195441	896	4e-95	82	AF327440_1 (AF327440) BTE-binding protein 4 [Homo sapiens]
302	gi14549656	731	5e-76	71	AF283891_1 (AF283891) dopamine receptor regulating factor [Mus musculus]
302	gi19919730	528	2e-52	46	AF490374_1 (AF490374) BTEB5 [Homo sapiens]
303	gi13159480	604	7e-62	100	(AX079973) Translation may initiate at the ATG codon at nucleotides 40-42 or the ATG at nucleotides 43-45 [Homo sapiens]
304	gi14164615	2143	0.0	100	AF310234_1 (AF310234) sialic acid binding immunoglobulin-like lectin 8 [Homo sapiens]
304	gi5541872	1295	e-141	69	(AJ130711) QA79 membrane protein, splice product airm-2 [Homo sapiens]
304	gi9837433	1320	e-144	96	AF287892_1 (AF287892) sialic acid binding immunoglobulin-like lectin 8 long splice variant [Homo sapiens]
305	gi11231111	437	2e-42	74	(AB051124) hypothetical protein [Macaca fascicularis]
306	gi4490795	1634	e-180	88	(AJ010341) cyclin-dependent kinase [Homo sapiens]
306	gi556651	1634	e-180	88	(X78342) PISSLRE [Homo sapiens]
306	gi8521453	1289	e-140	86	(L33264) CDC2-related protein kinase [Homo sapiens]
307	gi13939849	1819	0.0	100	(AX113671) chemokine receptor (CCX CKR) [Homo sapiens]
307	gi7274392	1819	0.0	100	(AF233281) CC chemokine receptor [Homo sapiens]
307	gi7363342	1819	0.0	100	AF193507_1 (AF193507) chemokine receptor [Homo sapiens]
308	gi24817412	877	3e-93	100	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
309	gi24817412	853	3e-90	99	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
310	gi24817412	264	9e-23	88	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
311	gi24817412	853	2e-90	99	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
312	gi17940754	3335	0.0	88	AF451975_1 (AF451975) cask-interacting protein 1 [Rattus norvegicus]
312	gi17940756	1441	e-157	54	AF451976_1 (AF451976) cask-interacting protein 2 [Homo sapiens]
312	gi17940758	3771	0.0	99	AF451977_1 (AF451977) cask-interacting

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					protein 1 [Homo sapiens]
313	gi1504040	4573	0.0	59	(D86983) similar to D.melanogaster peroxidasin(U11052) [Homo sapiens]
313	gi6273399	4573	0.0	59	AF200348_1 (AF200348) melanoma-associated antigen MG50 [Homo sapiens]
313	gi7292259	2604	0.0	38	(AE003475) CG12002-PA [Drosophila melanogaster]
314	gi28204826	2271	0.0	46	(BC046363) zinc-finger protein AY163807 [Homo sapiens]
314	gi6176338	4027	0.0	99	AF188530_1 (AF188530) ubiquitous tetratricopeptide containing protein RoXaN [Homo sapiens]
314	gi6562060	5211	0.0	98	(AL035659) dJ979N1.1 (dJ979N1.1) [Homo sapiens]
315	gi12654511	1843	0.0	88	AAH01085 (BC001085) ATP-dependant interferon response protein 1 [Homo sapiens]
315	gi14043167	1843	0.0	88	AAH07571 (BC007571) ATP-dependant interferon response protein 1 [Homo sapiens]
315	gi15079904	1843	0.0	88	AAH11746 (BC011746) ATP-dependant interferon response protein 1 [Homo sapiens]
316	gi7546797	2721	0.0	92	AF195833_1 (AF195833) cell adhesion molecule nectin-3 alpha [Mus musculus]
316	gi7546801	1794	0.0	93	AF195835_1 (AF195835) cell adhesion molecule nectin-3 gamma [Mus musculus]
316	gi9716665	2901	0.0	100	(AF282874) nectin 3; PRR3 [Homo sapiens]
317	gi16306735	1258	e-137	100	AAH01549 (BC001549) emopamil-binding protein (sterol isomerase) [Homo sapiens]
317	gi16306768	1258	e-137	100	AAH01572 (BC001572) emopamil-binding protein (sterol isomerase) [Homo sapiens]
317	gi28277024	1258	e-137	100	(BC046501) emopamil binding protein (sterol isomerase) [Homo sapiens]
318	gi21429160	153	6e-10	50	(AY119645) RE44650p [Drosophila melanogaster]
318	gi7296222	153	6e-10	50	(AE003590) CG11562-PA [Drosophila melanogaster]
319	gi10178883	3179	0.0	100	(AJ279016) chondrocyte expressed protein 68 kDa [Homo sapiens]
319	gi19171211	3367	0.0	100	(AJ421515) CRTAC1-B protein [Homo sapiens]
319	gi9368807	3179	0.0	100	(AJ276171) ASPIC [Homo sapiens]
320	gi16041826	984	e-105	68	AAH15803 (BC015803) interferon regulatory factor 2 [Homo sapiens]
320	gi19387294	960	e-102	65	AF480857_1 (AF480857) interferon regulatory factor 2 [Sigmodon hispidus]
320	gi33967	970	e-104	68	(X15949) interferon regulatory factor-2 (AA 1-349) [Homo sapiens]
321	gi10444285	1649	0.0	100	(AF290204) blood group carrier molecule DOK1 [Homo sapiens]
321	gi20385811	1649	0.0	100	(AF382213) Dombrock blood group carrier molecule [Homo sapiens]
321	gi20385818	1644	0.0	99	(AF382216) Dombrock blood group carrier molecule [Homo sapiens]
322	gi15077418	1385	e-151	100	AF326778_1 (AF326778) gastric cancer multidrug resistance-associated protein [Homo sapiens]
322	gi18535616	5262	0.0	90	(AY074490) EEG1L [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
322	gi18535618	1371	e-149	100	(AY074491) EEG1S [Homo sapiens]
323	gi15341958	147	9e-09	33	AAH13172 (BC013172) Similar to DKFZP564L0862 protein [Homo sapiens]
323	gi15420873	615	5e-63	97	AF398968_1 (AF398968) ankyrin repeat-containing SOCS box protein 7 [Mus musculus]
323	gi18031947	145	2e-08	34	(AY057053) SOCS box protein ASB-5 [Homo sapiens]
324	gi13477335	964	e-103	100	AAH05143 (BC005143) vitamin A responsive; cytoskeleton related [Homo sapiens]
324	gi18088541	964	e-103	100	AAH20797 (BC020797) vitamin A responsive; cytoskeleton related [Homo sapiens]
324	gi21217445	964	e-103	100	(AY102608) JWA protein [Homo sapiens]
325	gi15779083	1138	e-123	91	AAH14609 (BC014609) Unknown (protein for MGC:26973) [Homo sapiens]
325	gi3342737	983	e-105	88	(AC005328) R26660_2, partial CDS [Homo sapiens]
325	gi3478640	154	4e-09	100	(AC005545) R26660_2, partial CDS [Homo sapiens]
326	gi12805563	556	7e-56	85	(BC002259) Similar to anaphase-promoting complex subunit 4 [Mus musculus]
326	gi19353519	921	3e-98	85	(BC024870) RIKEN cDNA 2610306D21 gene [Mus musculus]
326	gi6180011	1074	e-116	100	AF191338_1 (AF191338) anaphase-promoting complex subunit 4 [Homo sapiens]
327	gi12597921	994	e-106	43	(U82982) GEC-3 [Cavia porcellus]
327	gi12718818	1017	e-109	45	(AB044284) sulfhydryl oxidase [Mus musculus]
327	gi22658418	1999	0.0	83	(BC030934) similar to quiescin [Mus musculus]
328	gi12804553	1592	e-176	100	AAH01689 (BC001689) carnitine/acylcarnitine translocase [Homo sapiens]
328	gi2765075	1592	e-176	100	(Y10319) carnitine carrier [Homo sapiens]
328	gi5851675	1582	e-174	99	(Y17775) carnitine/acylcarnitine translocase [Homo sapiens]
329	gi14602799	1302	e-142	92	AAH09907 (BC009907) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
329	gi15215451	1302	e-142	92	AAH12819 (BC012819) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
329	gi38522	1305	e-142	92	(Z21507) human elongation factor-1-delta [Homo sapiens]
330	gi14124972	860	5e-91	84	AAH08012 (BC008012) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
330	gi14602799	860	5e-91	84	AAH09907 (BC009907) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
330	gi15215451	860	5e-91	84	AAH12819 (BC012819) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
331	gi178257	1064	e-115	99	(M13692) alpha-1 acid glycoprotein

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					precursor [Homo sapiens]
331	gi20070760	1068	e-115	100	(BC026238) orosomucoid 1 [Homo sapiens]
331	gi757907	1064	e-115	99	(X02544) alpha1-acid glycoprotein [Homo sapiens]
332	gi17061809	593	2e-60	100	(AY040090) C21orf15 protein [Homo sapiens]
333	gi203699	565	2e-57	100	(K00750) cytochrome c [Rattus norvegicus]
333	gi21706378	565	2e-57	100	(BC034363) cytochrome c, somatic [Mus musculus]
333	gi50619	565	2e-57	100	(X01756) cytochrome c [Mus musculus]
334	gi15418732	2290	0.0	99	(AY008445) STAMP1 [Homo sapiens]
334	gi18677151	1311	e-143	57	(AF238865) tumor suppressor pHyde [Rattus norvegicus]
334	gi22655488	2284	0.0	99	AF455138_1 (AF455138) six-transmembrane epithelial antigen of prostate 2 [Homo sapiens]
335	gi11545707	138	4e-08	100	(AY009128) ISCU2 [Homo sapiens]
335	gi15080288	138	4e-08	100	AAH11906 (BC011906) Unknown (protein for MGC:20315) [Homo sapiens]
335	gi20381021	125	1e-06	93	(BC028800) RIKEN cDNA 2310020H20 gene [Mus musculus]
336	gi17224904	1952	0.0	43	AF317839_1 (AF317839) immunoglobulin superfamily member 9 [Mus musculus]
336	gi20988778	1910	0.0	42	(BC030141) Similar to immunoglobulin superfamily, member 9 [Homo sapiens]
336	gi25955616	1942	0.0	42	(BC040281) immunoglobulin superfamily, member 9 [Mus musculus]
337	gi26340432	1880	0.0	89	(AK049696) unnamed protein product [Mus musculus]
337	gi26352762	1880	0.0	89	(AK087811) unnamed protein product [Mus musculus]
337	gi5459205	2058	0.0	100	(AL031431) dJ462O23.2 (novel protein) [Homo sapiens]
338	gi17016967	5677	0.0	100	AF435011_1 (AF435011) NUANCE [Homo sapiens]
338	gi17861384	5677	0.0	100	(AY061759) nesprin-2 gamma [Homo sapiens]
338	gi24417711	5677	0.0	100	(AF495911) nesprin-2 [Homo sapiens]
339	gi14248997	2239	0.0	97	AF376725_1 (AF376725) lung seven transmembrane receptor 1 [Homo sapiens]
339	gi14248999	916	3e-97	47	AF376726_1 (AF376726) lung seven transmembrane receptor 2 [Mus musculus]
339	gi7291031	765	1e-79	50	(AE003446) CG12121-PA [Drosophila melanogaster]
340	gi14789614	1401	e-153	70	AAH10743 (BC010743) Similar to CGI-45 protein [Homo sapiens]
340	gi23271651	1692	0.0	99	(BC024094) Similar to CGI-45 protein [Mus musculus]
340	gi4929559	1385	e-151	71	AF151803_1 (AF151803) CGI-45 protein [Homo sapiens]
341	gi1542939	2087	0.0	54	(Y07903) transmembrane protein tMDC I [Rattus norvegicus]
341	gi1666651	2074	0.0	54	(X64227) Cyritestin [Mus musculus]
341	gi535017	3422	0.0	86	(X76637) tMDC I [Macaca fascicularis]
342	gi212451	182	6e-12	20	(M93676) nonmuscle myosin heavy chain [Gallus gallus]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
342	gi212452	182	6e-12	20	(M93676) nonmuscle myosin heavy chain [Gallus gallus]
342	gi641958	182	6e-12	20	(M69181) non-muscle myosin B [Homo sapiens]
343	gi211499	431	2e-41	43	(K01702) HMW/LMW collagen subunit precursor [Gallus gallus]
343	gi22652113	1065	e-115	98	AF406780_1 (AF406780) alpha 1 type XXII collagen [Homo sapiens]
343	gi298642	418	8e-40	46	(S57132) type XVI collagen alpha 1 chain; alpha 1 (XVI) [Homo sapiens]
344	gi1817733	4685	0.0	92	(U63834) KIT protein [Homo sapiens]
344	gi259336	4685	0.0	92	(S48745) mast/stem cell growth factor receptor [human]
344	gi34085	4685	0.0	92	(X06182) protein p145-ckit (AA 1 - 976) [Homo sapiens]
345	gi15217067	1376	e-151	96	AF400436_1 (AF400436) stem cell factor isoform 1 [Homo sapiens]
345	gi1827477	1195	e-130	84	(D50833) stem cell factor [Felis catus]
345	gi337934	1376	e-151	96	(M59964) stem cell factor [Homo sapiens]
346	gi19387136	3508	0.0	99	AF479748_1 (AF479748) PYRIN-containing APAF1-like protein 5 [Homo sapiens]
346	gi202806	1566	e-172	67	(M85183) vasopressin receptor [Rattus norvegicus]
346	gi21410402	1408	e-154	64	(BC031139) expressed sequence AI504961 [Mus musculus]
347	gi19387136	4563	0.0	99	AF479748_1 (AF479748) PYRIN-containing APAF1-like protein 5 [Homo sapiens]
347	gi202806	1566	e-172	67	(M85183) vasopressin receptor [Rattus norvegicus]
347	gi21410402	1408	e-154	64	(BC031139) expressed sequence AI504961 [Mus musculus]
348	gi17512442	601	2e-60	50	(BC019180) ficolin A [Mus musculus]
348	gi27085383	605	5e-61	54	(AY173052) microfibril-associated glycoprotein 4 [Bos taurus]
348	gi790817	661	2e-67	55	(L38486) microfibril-associated glycoprotein 4 [Homo sapiens]
349	gi17512442	601	1e-60	50	(BC019180) ficolin A [Mus musculus]
349	gi27085383	605	4e-61	54	(AY173052) microfibril-associated glycoprotein 4 [Bos taurus]
349	gi790817	661	1e-67	55	(L38486) microfibril-associated glycoprotein 4 [Homo sapiens]
350	gi11877276	533	8e-53	31	(AL121756) dJ726C3.5 (ortholog of potential ligand_binding protein RY2G5 (Rat)) [Homo sapiens]
350	gi21667214	2286	0.0	100	AF465767_1 (AF465767) bactericidal/permeability-increasing protein-like 3 [Homo sapiens]
350	gi57732	573	2e-57	33	(X60660) potential ligand-binding protein [Rattus rattus]
351	gi13183327	2363	0.0	100	AF274714_1 (AF274714) oxysterol-binding protein-related protein [Homo sapiens]
351	gi17529997	2351	0.0	99	AF392449_1 (AF392449) oxysterol-binding protein-like protein OSBPL1A [Homo sapiens]
351	gi17529999	2358	0.0	99	AF392450_1 (AF392450) oxysterol-binding protein-like protein OSBPL1B [Homo

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					sapiens]
352	gi21425644	229	1e-16	39	(AJ318215) putative E3 ubiquitin ligase [Homo sapiens]
352	gi27263233	229	1e-16	39	(AY145132) p53-associated parkin-like cytoplasmic protein [Homo sapiens]
352	gi559707	242	3e-18	41	(D38548) The ha0936 gene product is novel. [Homo sapiens]
353	gi13274524	1462	e-161	94	AF329839_1 (AF329839) complement-c1q tumor necrosis factor-related protein [Homo sapiens]
353	gi18381163	1462	e-161	94	AAH22187 (BC022187) complement-c1q tumor necrosis factor-related protein 7 [Homo sapiens]
353	gi18645144	1462	e-161	94	(BC024015) C1q and tumor necrosis factor related protein 7 [Homo sapiens]
354	gi23273642	695	3e-72	100	(BC036302) Similar to lymphocyte antigen 6 complex, locus G6C [Homo sapiens]
354	gi4337100	695	3e-72	100	AAD18076 (AF129756) G6c [Homo sapiens]
354	gi5304878	695	3e-72	100	(AJ012008) Ly6-C protein [Homo sapiens]
355	gi10198115	2760	0.0	100	AF279890_1 (AF279890) 2P domain potassium channel TREK2 [Homo sapiens]
355	gi19701864	2760	0.0	100	(AX393903) ORF of human TREK2 cDNA [Homo sapiens]
355	gi19716292	2690	0.0	99	AF385400_1 (AF385400) potassium channel TREK2 splice variant c [Homo sapiens]
356	gi10198115	2697	0.0	100	AF279890_1 (AF279890) 2P domain potassium channel TREK2 [Homo sapiens]
356	gi19701864	2697	0.0	100	(AX393903) ORF of human TREK2 cDNA [Homo sapiens]
356	gi19716292	2788	0.0	99	AF385400_1 (AF385400) potassium channel TREK2 splice variant c [Homo sapiens]
357	gi177870	2767	0.0	40	(M11313) alpha-2-macroglobulin precursor [Homo sapiens]
357	gi25303946	2767	0.0	40	(BC040071) alpha-2-macroglobulin [Homo sapiens]
357	gi579592	2761	0.0	40	(A21185) alpha 2-macroglobulin 690-730 [Homo sapiens]
358	gi1405744	2294	0.0	99	(X63963) Pax-6 (paired box containing gene) [Mus musculus]
358	gi18138028	2289	0.0	99	(Y19196) paired box protein [Mus musculus]
358	gi18138034	2294	0.0	99	(Y19199) paired box protein [Mus musculus]
359	gi27530341	592	6e-60	42	(AB016429) collectin-L1 [Mus musculus]
359	gi415939	309	4e-27	32	(X75911) lung surfactant protein D [Bos taurus]
359	gi5162875	612	3e-62	42	(AB002631) collectin 34 [Homo sapiens]
360	gi177179	597	2e-60	41	(M60832) alpha-2 type VIII collagen [Homo sapiens]
360	gi18496364	728	1e-75	46	(AB067770) otolin-1 [Oncorhynchus keta]
360	gi18676606	614	2e-62	41	(AK074129) FLJ00201 protein [Homo sapiens]
361	gi3228237	791	3e-83	69	(AJ006692) ultra high sulfer keratin [Homo sapiens]
361	gi32472	783	3e-82	76	(X63755) high-sulpher keratin [Homo sapiens]
361	gi34079	772	5e-81	76	(X55293) ultra high-sulphur keratin protein [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
362	gi200962	823	8e-87	66	(M37759) serine 1 ultra high sulfur protein [Mus musculus]
362	gi3228237	872	2e-92	73	(AJ006692) ultra high sulfer keratin [Homo sapiens]
362	gi32472	724	2e-75	69	(X63755) high-sulpher keratin [Homo sapiens]
363	gi15718478	561	4e-56	47	(AF257472) transmembrane protein MT75 [Homo sapiens]
363	gi17979839	575	9e-58	49	(AF311699) c-type lectin protein MT75 [Mus musculus]
363	gi3790610	1551	e-171	83	(AF093673) layilin [Cricetulus griseus]
365	gi12654511	2154	0.0	100	AAH01085 (BC001085) ATP-dependant interferon response protein 1 [Homo sapiens]
365	gi14043167	2154	0.0	100	AAH07571 (BC007571) ATP-dependant interferon response protein 1 [Homo sapiens]
365	gi15079904	2154	0.0	100	AAH11746 (BC011746) ATP-dependant interferon response protein 1 [Homo sapiens]
366	gi12654511	1843	0.0	88	AAH01085 (BC001085) ATP-dependant interferon response protein 1 [Homo sapiens]
366	gi14043167	1843	0.0	88	AAH07571 (BC007571) ATP-dependant interferon response protein 1 [Homo sapiens]
366	gi15079904	1843	0.0	88	AAH11746 (BC011746) ATP-dependant interferon response protein 1 [Homo sapiens]
368	gi10435784	1011	e-108	100	(AK023755) unnamed protein product [Homo sapiens]
368	gi27451951	1005	e-108	99	(AF534824) TREM-like transcript 2 [Homo sapiens]
369	gi10566471	1375	e-150	99	(AB044560) Gliacolin [Mus musculus]
369	gi14278927	1375	e-150	99	(AB045983) gliacolin [Mus musculus]
369	gi27817288	1152	e-125	86	(AL672065) SI:dZ63M2.2 (novel protein similar to gliacolin) [Danio rerio]
370	gi20071655	375	1e-34	37	(BC027426) cellular repressor of E1A-stimulated genes [Mus musculus]
370	gi24371079	1547	e-170	100	(AB046109) CREG2 [Homo sapiens]
370	gi24371081	1286	e-140	83	(AB046110) CREG2 [Mus musculus]
371	gi11090860	168	8e-11	24	AF251509_1 (AF251509) leukocyte-associated Ig-like receptor 1C isoform; LAIR-1C [Homo sapiens]
371	gi16930383	172	3e-11	38	AF383169_1 (AF383169) leukocyte immunoglobulin-like receptor e [Pan troglodytes]
371	gi6563042	179	4e-12	24	AF109683_1 (AF109683) leukocyte-associated Ig-like receptor 1b [Homo sapiens]
372	gi11120574	260	3e-22	100	AF309653_1 (AF309653) CD20/Fc-epsilon-RI-beta family member 4 [Homo sapiens]
372	gi18028930	260	3e-22	100	AF350501_1 (AF350501) four-span transmembrane protein 2 [Homo sapiens]
372	gi18089082	260	3e-22	100	AAH20673 (BC020673) membrane-spanning 4-domains, subfamily A, member 7 [Homo sapiens]
373	gi17391109	229	1e-18	82	AAH18471 (BC018471) Similar to nitrogen fixation gene 1 (S. cerevisiae, homolog) [Homo sapiens]
373	gi21595759	223	7e-18	82	(BC032569) similar to HC6 [Homo sapiens]
373	gi6690252	236	2e-19	84	AF090944_1 (AF090944) PRO0663 [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
374	gi1684833	3087	0.0	93	(U77667) tyrosine kinase [Mus musculus]
374	gi20987557	3102	0.0	93	(BC029727) zeta-chain (TCR) associated protein kinase (70kD) [Mus musculus]
374	gi436480	3084	0.0	93	(U04379) ZAP-70 [Mus musculus]
375	gi12002311	2780	0.0	100	AF142573_1 (AF142573) putative secretory protein precursor [Homo sapiens]
375	gi13241974	2780	0.0	100	AF329197_1 (AF329197) CocoaCrisp [Homo sapiens]
375	gi18088175	2780	0.0	100	AAH20514 (BC020514) CocoaCrisp [Homo sapiens]
376	gi15559680	1803	0.0	100	AAH14195 (BC014195) hypothetical protein FLJ21172 [Homo sapiens]
376	gi18447566	185	1e-12	27	(AY075537) RH08992p [Drosophila melanogaster]
376	gi22832309	185	1e-12	27	(AE003500) CG15916-PA [Drosophila melanogaster]
377	gi20988290	781	4e-82	100	(BC029889) similar to evidence:NAS~putative~unclassifiable [Homo sapiens]
377	gi27899963	740	2e-77	97	(AX588217) unnamed protein product [Homo sapiens]
377	gi27899965	751	1e-78	99	(AX588218) unnamed protein product [Homo sapiens]
378	gi20988290	351	6e-33	98	(BC029889) similar to evidence:NAS~putative~unclassifiable [Homo sapiens]
378	gi27899963	317	5e-29	95	(AX588217) unnamed protein product [Homo sapiens]
378	gi27899965	321	2e-29	97	(AX588218) unnamed protein product [Homo sapiens]
379	gi21594969	472	1e-46	100	(BC031610) membrane-spanning 4-domains, subfamily A, member 12 4-domains, subfamily A, member 7 [Homo sapiens]
380	gi16041675	575	2e-58	100	AAH15704 (BC015704) joined to JAZF1 [Homo sapiens]
380	gi23093099	139	7e-08	36	AE003515_36 (AE003515) CG8013-PB [Drosophila melanogaster]
380	gi23093100	139	7e-08	36	(AE003515) CG8013-PA [Drosophila melanogaster]
381	gi14669826	1787	0.0	90	(AB057731) lipoic acid synthase [Mus musculus]
381	gi23958222	1975	0.0	99	(BC023635) Similar to lipoic acid synthetase [Homo sapiens]
381	gi7296306	1241	e-135	67	(AE003591) CG5231-PA [Drosophila melanogaster]
382	gi16118499	485	1e-47	58	AF397035_9 (AF397035) G7d [Mus musculus]
382	gi16118508	485	1e-47	58	AF397036_9 (AF397036) G7d [Mus musculus]
382	gi4529898	734	1e-76	82	(AF134726) NG23 [Homo sapiens]
383	gi11066090	1188	e-128	85	AF195192_1 (AF195192) matrix metalloprotease MMP-27 [Homo sapiens]
383	gi12006364	1121	e-121	81	AF281673_1 (AF281673) matrix metalloproteinase-27 [Tupaia belangeri]
383	gi180618	923	5e-98	63	(J05556) neutrophil collagenase [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
384	gi24251209	4600	0.0	100	(AY149237) collagen XXVII proalpha 1 chain precursor; preproprotein [Homo sapiens]
384	gi28172191	4147	0.0	89	(AL683828) bM340H1.1 (novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein) [Mus musculus]
384	gi28204656	4147	0.0	89	(AY167568) collagen type XXVII proalpha 1 chain [Mus musculus]
385	gi15215576	2580	0.0	76	(AY050249) BMP-2 inducible kinase [Mus musculus]
385	gi23271902	783	1e-81	98	(BC036021) Similar to Bmp2-inducible kinase [Homo sapiens]
385	gi3970852	1132	e-122	100	(AB015331) HRIHFB2017 [Homo sapiens]
387	gi14043517	1539	e-169	100	AAH07744 (BC007744) Unknown (protein for MGC:13286) [Homo sapiens]
387	gi6682314	328	3e-29	33	(AL022072) conserved protein; possibly mitochondrial protein synthesis; DUF28 domain [Schizosaccharomyces pombe]
387	gi6690225	653	6e-67	99	AF090929_2 (AF090929) PRO0477p [Homo sapiens]
388	gi10437569	354	1e-32	70	(AK025116) unnamed protein product [Homo sapiens]
388	gi21748687	351	3e-32	69	(AK090511) unnamed protein product [Homo sapiens]
388	gi7020625	331	7e-30	62	(AK000496) unnamed protein product [Homo sapiens]
389	gi12843048	343	3e-31	72	(AK008696) unnamed protein product [Mus musculus]
389	gi26329371	435	7e-42	59	(AK033677) unnamed protein product [Mus musculus]
389	gi26354052	435	7e-42	59	(AK088927) unnamed protein product [Mus musculus]
390	gi12843048	343	4e-31	72	(AK008696) unnamed protein product [Mus musculus]
390	gi26329371	435	8e-42	59	(AK033677) unnamed protein product [Mus musculus]
390	gi26354052	436	6e-42	55	(AK088927) unnamed protein product [Mus musculus]
392	gi17426496	808	9e-85	50	(AL590222) bA159L8.1 (putative purinergic receptor (FKSG79)) [Homo sapiens]
392	gi2104787	1792	0.0	100	(AF000545) putative purinergic receptor P2Y10 [Homo sapiens]
392	gi4455508	1792	0.0	100	(Z82200) dJ333E23.1 (7 transmembrane receptor) [Homo sapiens]
393	gi17426496	808	8e-85	50	(AL590222) bA159L8.1 (putative purinergic receptor (FKSG79)) [Homo sapiens]
393	gi2104787	1792	0.0	100	(AF000545) putative purinergic receptor P2Y10 [Homo sapiens]
393	gi4455508	1792	0.0	100	(Z82200) dJ333E23.1 (7 transmembrane receptor) [Homo sapiens]
394	gi14272704	1428	e-157	99	(AX136297) unnamed protein product [Homo sapiens]
394	gi19575509	1440	e-158	100	(AX380599) unnamed protein product [Homo sapiens]
394	gi19575655	1440	e-158	100	(AX380745) unnamed protein product [Homo

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					sapiens]
395	gi11127646	2253	0.0	100	AF149825_1 (AF149825) PACSIN3 [Homo sapiens]
395	gi13539688	2253	0.0	100	AF242530_1 (AF242530) protein kinase C and casein kinase substrate 3 [Homo sapiens]
395	gi14043958	2253	0.0	100	AAH07914 (BC007914) protein kinase C and casein kinase substrate in neurons 3 [Homo sapiens]
396	gi12805195	2370	0.0	90	(BC002056) heat shock protein, 70 kDa 4 [Mus musculus]
396	gi6563208	2554	0.0	99	AF112210_1 (AF112210) heat shock protein hsp70-related protein [Homo sapiens]
396	gi7672784	2557	0.0	99	AF143723_1 (AF143723) heat shock protein HSP60 [Homo sapiens]
397	gi178677	717	4e-74	36	(M17303) carcinoembryonic antigen precursor [Homo sapiens]
397	gi180223	717	4e-74	36	(M29540) carcinoembryonic antigen [Homo sapiens]
397	gi21961634	720	2e-74	36	(BC034671) carcinoembryonic antigen-related cell adhesion molecule 5 [Homo sapiens]
398	gi178677	462	1e-44	32	(M17303) carcinoembryonic antigen precursor [Homo sapiens]
398	gi180211	462	1e-44	32	(M59710) carcinoembryonic antigen [Homo sapiens]
398	gi21961634	465	6e-45	32	(BC034671) carcinoembryonic antigen-related cell adhesion molecule 5 [Homo sapiens]
399	gi178677	442	3e-42	33	(M17303) carcinoembryonic antigen precursor [Homo sapiens]
399	gi180211	442	3e-42	33	(M59710) carcinoembryonic antigen [Homo sapiens]
399	gi21961634	445	2e-42	34	(BC034671) carcinoembryonic antigen-related cell adhesion molecule 5 [Homo sapiens]
400	gi1061159	1277	e-139	37	(X87205) testicular Metalloprotease-like, Disintegrin-like, Cysteine-rich protein IVa [Macaca fascicularis]
400	gi26278978	2199	0.0	54	(AY158688) ADAM4 [Mus musculus]
400	gi965014	1407	e-154	53	(U22058) ADAM 4 protein precursor [Mus musculus]
401	gi1061161	496	1e-48	42	(X87206) testicular Metalloprotease-like, Disintegrin-like, Cysteine-rich protein IVb [Macaca fascicularis]
401	gi1061163	498	6e-49	43	(X87207) testicular Metalloprotease-like, Disintegrin-like, Cysteine-rich protein IVc [Macaca fascicularis]
401	gi26278978	777	3e-81	53	(AY158688) ADAM4 [Mus musculus]
402	gi11493443	2151	0.0	99	AF130117_27 (AF130068) PRO2209 [Homo sapiens]
402	gi177829	2151	0.0	99	(K01396) alpha-1-antitrypsin [Homo sapiens]
402	gi28966	2151	0.0	99	(X01683) alpha 1-antitrypsin [Homo sapiens]
403	gi21595832	2531	0.0	71	(BC032753) Kruppel-type zinc finger (C2H2) [Homo sapiens]
403	gi4519270	2531	0.0	71	(AB011414) Kruppel-type zinc finger protein [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
403	gi6467202	3321	0.0	99	(AB021642) gonadotropin inducible transcription repressor-2 [Homo sapiens]
404	gi12804197	1084	e-117	80	AAH02956 (BC002956) ClpP (caseinolytic protease, ATP-dependent, proteolytic subunit, E. coli) homolog [Homo sapiens]
404	gi12805083	817	4e-86	66	(BC001998) caseinolytic protease, ATP-dependent, (E. coli) proteolytic subunit homolog [Mus musculus]
404	gi963048	1084	e-117	80	(Z50853) CLPP [Homo sapiens]
405	gi180227	560	2e-56	80	(L00692) carcinoembryonic antigen [Homo sapiens]
405	gi219535	564	6e-57	81	(D90277) nonspecific cross-reacting antigen [Homo sapiens]
405	gi3851200	404	2e-38	60	(AC005955) CGM7_HUMAN [Homo sapiens]
406	gi15214636	1319	e-144	100	AAH12444 (BC012444) Similar to chloride intracellular channel 4 [Homo sapiens]
406	gi28204905	1304	e-142	98	(BC046384) chloride intracellular channel 4 (mitochondrial) [Mus musculus]
406	gi5052202	1305	e-142	99	AF097330_1 (AF097330) H1 chloride channel; p64H1; CLIC4 [Homo sapiens]
408	gi17389410	1439	e-158	100	AAH17745 (BC017745) Similar to nuclear fragile X mental retardation protein interacting protein 1 [Homo sapiens]
408	gi6525071	2611	0.0	97	(AF159548) nuclear FMRP interacting protein 1 [Homo sapiens]
408	gi6525073	1806	0.0	69	(AF159549) nuclear FMRP interacting protein 1 [Mus musculus]
409	gi21619491	473	2e-46	69	(BC031566) similar to expressed sequence AW049604 [Homo sapiens]
409	gi24658290	252	7e-21	51	(BC039396) Similar to expressed sequence AW049604 [Homo sapiens]
409	gi6572294	252	7e-21	51	(AL096843) bA262A13.1 (novel protein) [Homo sapiens]
410	gi14336713	3060	0.0	100	AE006464_13 (AE006464) possible G-protein receptor [Homo sapiens]
410	gi22478039	2261	0.0	99	(BC036680) Similar to expressed sequence AW322056 [Homo sapiens]
410	gi5912459	1110	e-119	100	(Z97653) c380A1.1 (novel protein) [Homo sapiens]
411	gi13625304	495	7e-49	59	AF293340_1 (AF293340) collagen-like Alzheimer amyloid plaque component precursor type I [Homo sapiens]
411	gi13649767	500	2e-49	57	AF315290_1 (AF315290) collagen-like Alzheimer amyloid plaque component precursor type I [Mus musculus]
411	gi22652221	889	1e-94	96	AF410792_1 (AF410792) alpha 1 type XXIII collagen [Mus musculus]
412	gi10998440	3167	0.0	69	AF276425_1 (AF276425) EGF-related protein SCUBE1 [Mus musculus]
412	gi25992504	3884	0.0	79	(AF525689) signal peptide-CUB-EGF-like domain containing protein 1 [Homo sapiens]
412	gi8052237	2916	0.0	58	(AJ400877) CEGP1 protein [Homo sapiens]
413	gi10998440	3151	0.0	69	AF276425_1 (AF276425) EGF-related protein SCUBE1 [Mus musculus]
413	gi25992504	3868	0.0	79	(AF525689) signal peptide-CUB-EGF-like

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
					domain containing protein 1 [Homo sapiens]
413	gi8052237	2898	0.0	58	(AJ400877) CEGP1 protein [Homo sapiens]
414	gi19354073	248	1e-20	68	(BC024666) cytochrome c oxidase, subunit VIc [Mus musculus]
414	gi203519	251	5e-21	68	(M27466) cytochrome c oxidase subunit VIc [Rattus norvegicus]
414	gi203710	251	5e-21	68	(M20153) cytochrome c oxidase subunit VIc [Rattus norvegicus]
415	gi15559697	157	2e-09	28	AAH14205 (BC014205) Similar to neural cell adhesion molecule 1 [Homo sapiens]
415	gi24620457	156	2e-09	26	(AY130758) 301KDa_2 protein [Caenorhabditis elegans]
415	gi61	158	1e-09	28	(X16451) calmodulin-independent adenylate cyclase [Bos taurus]
416	gi21432076	641	1e-65	58	(BC032975) RIKEN cDNA 4932438H23 gene [Mus musculus]
416	gi23342580	983	e-105	91	(AX497196) unnamed protein product [Homo sapiens]
416	gi8118227	1311	e-143	100	(AF231922) C21orf62 protein [Homo sapiens]
417	gi19569541	353	8e-32	42	AF485812_1 (AF485812) Fc gamma receptor I [Macaca fascicularis]
417	gi21619686	351	1e-31	41	(BC032634) Fc fragment of IgG, high affinity Ia, receptor for (CD64) [Homo sapiens]
417	gi31332	354	6e-32	41	(X14356) FcRI (AA 1-374) [Homo sapiens]
418	gi21205864	1591	e-175	100	AF385435_1 (AF385435) T-cell activation protein phosphatase 2C; TA-PP2C [Homo sapiens]
418	gi21464366	758	4e-79	52	(AY121659) RE06653p [Drosophila melanogaster]
418	gi7292094	758	4e-79	52	(AE003472) CG12091-PA [Drosophila melanogaster]
419	gi190568	1476	e-162	87	(M94890) pregnancy-specific beta-1 glycoprotein [Homo sapiens]
419	gi190647	1470	e-161	85	(M69245) pregnancy-specific beta-1-glycoprotein [Homo sapiens]
419	gi609318	1475	e-162	88	(U18469) pregnancy-specific beta 1-glycoprotein 4 precursor [Homo sapiens]
420	gi24412825	272	3e-23	100	(AL109928) dJ551D2.1.3 (Cadherin-like 26, variant 3) [Homo sapiens]
420	gi7981304	575	2e-58	84	(AL109928) dJ551D2.1.2 (Cadherin-like 26, variant 2) [Homo sapiens]
420	gi9622236	272	3e-23	100	AF169690_1 (AF169690) cadherin-like protein VR20 [Homo sapiens]
421	gi12833891	465	2e-45	55	(AK003305) unnamed protein product [Mus musculus]
421	gi23273040	991	e-106	99	(BC035810) Unknown (protein for IMAGE:5754421) [Homo sapiens]
421	gi24817754	465	2e-45	55	(AB095543) high density lipoprotein binding protein 1 [Mus musculus]
423	gi13241972	232	5e-18	33	AF329196_1 (AF329196) SugarCrisp [Mus musculus]
423	gi9558454	253	2e-20	33	(AB046537) cysteine-rich protease inhibitor [Mus musculus]
423	gi9558479	253	2e-20	33	(AB046539) cysteine-rich protease inhibitor [Mus musculus]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
424	gi13375149	961	e-103	100	(AL109964) dJ1118M15.2 (Novel protein) [Homo sapiens]
424	gi5442036	142	7e-08	31	AF109126_1 (AF109126) stromal cell-derived receptor-1 beta [Homo sapiens]
424	gi7259265	314	8e-28	50	(AB030198) contains transmembrane (TM) region [Mus musculus]
425	gi18480302	1007	e-108	79	(AY073502) olfactory receptor MOR262-10 [Mus musculus]
425	gi28279464	1008	e-108	79	(BC046311) olfactory receptor 70 [Mus musculus]
425	gi5869927	950	e-101	76	(AJ133430) olfactory receptor [Mus musculus]
426	gi21622561	1086	e-117	100	(AJ315545) LY6G5B protein [Homo sapiens]
426	gi5701854	794	2e-83	100	(AJ245417) LY6G5b protein [Homo sapiens]
426	gi6137324	789	7e-83	99	AF129756_1 (AF129756) G5b [Homo sapiens]
427	gi12652993	491	7e-49	100	AAH00257 (BC000257) Unknown (protein for IMAGE:3357862) [Homo sapiens]
427	gi14043883	491	7e-49	100	AAH07882 (BC007882) Similar to RIKEN cDNA 0610012G03 gene [Homo sapiens]
427	gi18204855	340	2e-31	75	(BC021536) Similar to RIKEN cDNA 0610012G03 gene [Mus musculus]
428	gi21432071	307	2e-27	65	(BC032982) Unknown (protein for MGC:41689) [Mus musculus]
429	gi13508539	162	4e-09	31	(AJ276961) CLASP2 [Mus musculus]
429	gi21064295	223	3e-16	31	(AY113372) LP02990p [Drosophila melanogaster]
429	gi7296250	223	3e-16	31	(AE003590) CG4648-PA [Drosophila melanogaster]
430	gi178991	1213	e-132	98	(M83751) arginine-rich protein [Homo sapiens]
430	gi27696986	706	3e-73	77	(BC043846) Similar to arginine-rich, mutated in early stage tumors [Xenopus laevis]
430	gi7300136	452	1e-43	54	(AE003713) CG7013-PA [Drosophila melanogaster]
431	gi17944240	169	8e-11	25	(AY070543) LD24657p [Drosophila melanogaster]
431	gi5020383	223	4e-17	32	(AF153450) juvenile hormone esterase binding protein [Manduca sexta]
431	gi7291887	169	8e-11	25	(AE003465) CG3776-PA [Drosophila melanogaster]
432	gi15862484	448	8e-44	96	(AX247850) unnamed protein product [Homo sapiens]
432	gi21619033	460	3e-45	88	(BC032306) Similar to RIKEN cDNA 2300005B03 gene [Homo sapiens]
432	gi28208164	533	1e-53	100	(AB081838) secreted Ly6/uPAR related protein 2 [Homo sapiens]
434	gi20521025	3343	0.0	100	(AB006623) No similarities to any reported proteins [Homo sapiens]
434	gi2706875	140	5e-07	25	(D85084) NCAM-180 [Cynops pyrrhogaster]
434	gi7768739	676	3e-69	30	(AP001745) human cDNA DKFZp586F0422, Accession No. AL050173 [Homo sapiens]
435	gi21542522	1052	e-113	45	(BC033024) AUT-like 1, cysteine endopeptidase (S. cerevisiae) [Homo sapiens]
435	gi27763975	2569	0.0	100	(AJ312332) APG4-D protein [Homo sapiens]
435	gi27763977	2181	0.0	86	(AJ312333) APG4-D protein [Mus musculus]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
436	gi190649	2009	0.0	87	(M93061) pregnancy-specific beta-1 glycoprotein [Homo sapiens]
436	gi300091	2009	0.0	87	(S59493) pregnancy-specific beta 1-glycoprotein; PSG [Homo sapiens]
436	gi904281	2008	0.0	87	(A23031) trophoblast membrane expressed protein [Homo sapiens]
437	gi15214951	1553	e-171	87	AAH12607 (BC012607) Similar to pregnancy specific beta-1-glycoprotein 5 [Homo sapiens]
437	gi190634	1534	e-169	86	(M73713) pregnancy-specific beta-1-glycoprotein 5 [Homo sapiens]
437	gi190638	1532	e-169	86	(M25384) fetal liver non-specific cross-reactive antigen-3 precursor protein [Homo sapiens]
438	gi13543533	1987	0.0	86	AAH05924 (BC005924) pregnancy specific beta-1-glycoprotein 3 [Homo sapiens]
438	gi180235	1899	0.0	86	(M37399) carcinoembryonic antigen SG5 [Homo sapiens]
438	gi904281	1899	0.0	86	(A23031) trophoblast membrane expressed protein [Homo sapiens]
439	gi13183078	1622	e-179	64	AF237652_1 (AF237652) a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 [Homo sapiens]
439	gi15099921	2352	0.0	95	AF176313_1 (AF176313) ADAM-TS related protein 1 [Homo sapiens]
439	gi20987759	2432	0.0	100	(BC030262) Similar to ADAMTS-like 1 [Homo sapiens]
440	gi13183078	2432	0.0	62	AF237652_1 (AF237652) a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 [Homo sapiens]
440	gi15099921	2907	0.0	99	AF176313_1 (AF176313) ADAM-TS related protein 1 [Homo sapiens]
440	gi20987759	2364	0.0	96	(BC030262) Similar to ADAMTS-like 1 [Homo sapiens]
441	gi13183078	2484	0.0	60	AF237652_1 (AF237652) a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 [Homo sapiens]
441	gi13625178	2343	0.0	100	AF251058_1 (AF251058) thrombospondin [Homo sapiens]
441	gi15099921	2798	0.0	99	AF176313_1 (AF176313) ADAM-TS related protein 1 [Homo sapiens]
442	gi15088529	124	3e-06	28	(AF319173) prostate stem cell antigen [Mus musculus]
442	gi1536902	560	9e-57	100	(X99977) ARS [Homo sapiens]
442	gi4218459	400	3e-38	69	(AJ132356) ARS component B precursor [Mus musculus]
443	gi21411513	658	5e-68	100	(BC031330) lymphocyte antigen 6 complex, locus D [Homo sapiens]
443	gi2739294	658	5e-68	100	(Y12642) E48 antigen [Homo sapiens]
443	gi887454	653	2e-67	99	(X82693) E48 antigen [Homo sapiens]
444	gi21411513	287	2e-25	96	(BC031330) lymphocyte antigen 6 complex, locus D [Homo sapiens]
444	gi2739294	287	2e-25	96	(Y12642) E48 antigen [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
444	gi887454	282	7e-25	94	(X82693) E48 antigen [Homo sapiens]
445	gi21428872	129	7e-06	25	(AY119501) GH11358p [Drosophila melanogaster]
445	gi21626538	129	7e-06	25	(AE003456) CG11170-PB [Drosophila melanogaster]
445	gi7291385	129	7e-06	25	(AE003456) CG11170-PA [Drosophila melanogaster]
446	gi13358942	3017	0.0	99	(AB056426) hypothetical protein [Macaca fascicularis]
446	gi13874489	2996	0.0	99	(AB060846) hypothetical protein [Macaca fascicularis]
446	gi26330992	2950	0.0	97	(AK035882) unnamed protein product [Mus musculus]
447	gi20258598	1742	0.0	100	(AY040542) sialic acid binding immunoglobulin-like lectin 6 [Homo sapiens]
447	gi2913995	1742	0.0	100	(D86358) CD33L1 [Homo sapiens]
447	gi2913997	1829	0.0	100	(D86359) CD33L2 [Homo sapiens]
448	gi1418928	7194	0.0	99	(Z74615) prepro-alpha1(I) collagen [Homo sapiens]
448	gi4755085	7197	0.0	99	(AF017178) pro alpha 1(I) collagen [Homo sapiens]
448	gi4960163	7105	0.0	98	AF153062_1 (AF153062) type I collagen pre-pro-alpha1(I) chain [Canis familiaris]
449	gi19068188	516	2e-51	64	(AY071842) IL-1F8 [Mus musculus]
449	gi6694394	818	2e-86	100	AF201833_1 (AF201833) FIL1 eta [Homo sapiens]
449	gi7769116	452	5e-44	94	AF200494_1 (AF200494) interleukin-1 homolog 2 [Homo sapiens]
450	gi15012124	278	8e-24	59	(BC010970) Similar to distal intestinal serine protease [Mus musculus]
450	gi26007900	278	8e-24	59	(BC040348) similar to distal intestinal serine protease [Mus musculus]
450	gi27370810	810	2e-85	100	(BC041609) Similar to distal intestinal serine protease [Homo sapiens]
451	gi15012124	1001	e-107	61	(BC010970) Similar to distal intestinal serine protease [Mus musculus]
451	gi26007900	1001	e-107	61	(BC040348) similar to distal intestinal serine protease [Mus musculus]
451	gi5921501	991	e-106	61	(AJ243866) distal intestinal serine protease [Mus musculus]
452	gi13938436	1017	e-109	100	AAH07359 (BC007359) Unknown (protein for IMAGE:3622437) [Homo sapiens]
452	gi19908462	798	1e-83	81	AF265232_1 (AF265232) rotatin [Mus musculus]
452	gi23271829	1657	0.0	83	(BC023916) Unknown (protein for IMAGE:5323200) [Mus musculus]
453	gi15029694	1954	0.0	58	(BC011061) procollagen, type VIII, alpha 1 [Mus musculus]
453	gi177179	3520	0.0	97	(M60832) alpha-2 type VIII collagen [Homo sapiens]
453	gi18676606	3953	0.0	100	(AK074129) FLJ00201 protein [Homo sapiens]
454	gi178991	148	5e-09	59	(M83751) arginine-rich protein [Homo sapiens]
454	gi27551197	410	2e-39	96	(AX573504) unnamed protein product [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
454	gi27696986	150	3e-09	43	(BC043846) Similar to arginine-rich, mutated in early stage tumors [Xenopus laevis]
455	gi21753515	130	7e-07	55	(AK094450) unnamed protein product [Homo sapiens]
456	gi1695690	142	2e-08	42	(D86232) Ly-6C variant [Mus musculus]
456	gi205250	144	1e-08	44	(M30690) Ly6C antigen [Rattus norvegicus]
456	gi52959	143	2e-08	41	(X04653) precursor polypeptide (AA -26 to 108) [Mus musculus]
457	gi11385997	1937	0.0	50	AF316985_1 (AF316985) toll-like receptor 1 [Mus musculus]
457	gi11528627	1932	0.0	50	(AY009154) toll-like receptor 1 [Mus musculus]
457	gi13447753	4277	0.0	100	AF296673_1 (AF296673) toll-like receptor 10 [Homo sapiens]
459	gi12406754	195	4e-14	73	(AX061647) unnamed protein product [Homo sapiens]
459	gi18378673	196	3e-14	76	AF462605_1 (AF462605) PATE [Homo sapiens]
460	gi1536902	204	2e-15	42	(X99977) ARS [Homo sapiens]
460	gi21411513	133	4e-07	37	(BC031330) lymphocyte antigen 6 complex, locus D [Homo sapiens]
460	gi4218459	219	4e-17	44	(AJ132356) ARS component B precursor [Mus musculus]
462	gi1542939	2050	0.0	52	(Y07903) transmembrane protein tMDC I [Rattus norvegicus]
462	gi1666651	2031	0.0	52	(X64227) Cyritestin [Mus musculus]
462	gi535017	3379	0.0	83	(X76637) tMDC I [Macaca fascicularis]
463	gi1542939	997	e-106	56	(Y07903) transmembrane protein tMDC I [Rattus norvegicus]
463	gi1666651	1032	e-111	57	(X64227) Cyritestin [Mus musculus]
463	gi535017	1517	e-167	83	(X76637) tMDC I [Macaca fascicularis]
464	gi531478	1487	e-163	76	(X77619) tMDC II [Macaca fascicularis]
464	gi965006	943	e-100	50	(U22060) ADAM 5 protein precursor [Cavia porcellus]
464	gi965016	844	6e-89	44	(U22059) ADAM 5 protein precursor [Mus musculus]
465	gi531478	1208	e-131	82	(X77619) tMDC II [Macaca fascicularis]
465	gi965006	804	3e-84	56	(U22060) ADAM 5 protein precursor [Cavia porcellus]
465	gi965016	678	1e-69	47	(U22059) ADAM 5 protein precursor [Mus musculus]
466	gi15779024	589	6e-60	53	AAH14588 (BC014588) Similar to acrosomal vesicle protein 1 [Homo sapiens]
466	gi338294	589	6e-60	53	(M82968) sperm protein 10 [Homo sapiens]
466	gi7705047	581	5e-59	53	(S65583) SP-10 [Homo sapiens]
467	gi15779024	741	2e-77	61	AAH14588 (BC014588) Similar to acrosomal vesicle protein 1 [Homo sapiens]
467	gi338292	771	6e-81	66	(M82967) sperm protein 10 [Homo sapiens]
467	gi338294	741	2e-77	61	(M82968) sperm protein 10 [Homo sapiens]
468	gi15779024	865	9e-92	69	AAH14588 (BC014588) Similar to acrosomal vesicle protein 1 [Homo sapiens]
468	gi338294	865	9e-92	69	(M82968) sperm protein 10 [Homo sapiens]
468	gi7705047	857	8e-91	68	(S65583) SP-10 [Homo sapiens]
469	gi15779024	746	5e-78	62	AAH14588 (BC014588) Similar to acrosomal vesicle protein 1 [Homo sapiens]
469	gi338294	746	5e-78	62	(M82968) sperm protein 10 [Homo sapiens]

TABLE 2A

SEQ ID	Hit ID	B score	P value	Percentage identity	Description
469	gi7705047	746	5e-78	62	(S65583) SP-10 [Homo sapiens]
470	gi15779024	459	6e-45	82	AAH14588 (BC014588) Similar to acrosomal vesicle protein 1 [Homo sapiens]
470	gi298489	464	2e-45	79	(S56458) SP-10 [Papio hamadryas] [Papio papio]
470	gi338292	468	5e-46	83	(M82967) sperm protein 10 [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
236	gi20198487	5657	0.0	99	AF441771_1 (AF441771) 182kDa tankyrase1-binding protein [Homo sapiens]
236	gi18676574	2747	0.0	99	(AK074113) FLJ00184 protein [Homo sapiens]
236	gi19684154	2695	0.0	67	(BC025943) Tnks1bp1 protein [Mus musculus]
237	gi40788181	6065	0.0	100	(AJ583821) ubiquitin specific proteinase 40 [Homo sapiens]
237	gi37361828	778	8e-81	54	(AY387057) LRRGT00071 [Rattus norvegicus]
237	gi10998129	437	3e-41	37	(AP002040) ubiquitin carboxyl-terminal hydrolase-like protein [Arabidopsis thaliana]
238	gi16506257	1652	0.0	99	AF329488_1 (AF329488) IFGP1 [Homo sapiens]
238	gi15528831	1640	0.0	99	(AY043464) Fc receptor-like protein 1 [Homo sapiens]
238	gi18140081	1640	0.0	99	AF459634_1 (AF459634) immunoglobulin superfamily receptor translocation associated 5 [Homo sapiens]
239	gi21104492	743	2e-78	100	(AB064665) OK/SW-CL.16 [Homo sapiens]
239	gi1372963	178	8e-13	68	(M85148) cytochrome oxidase subunit III [Macaca mulatta]
240	gi8745547	528	2e-53	100	AF268037_1 (AF268037) C8ORF4 protein [Homo sapiens]
240	gi18203818	528	2e-53	100	(BC021672) Chromosome 8 open reading frame 4 [Homo sapiens]
240	gi27503415	119	6e-06	49	(BC042280) LOC398479 protein [Xenopus laevis]
241	gi30584529	1128	e-122	100	(BT007845) Homo sapiens chorionic somatomammotropin hormone 1 (placental lactogen) [synthetic construct]
241	gi30584141	1128	e-122	100	(BT007651) Homo sapiens chorionic somatomammotropin hormone 1 (placental lactogen) [synthetic construct]
241	gi190034	1128	e-122	100	(J00118) placental lactogen [Homo sapiens]
242	gi14575679	3662	0.0	96	AF156100_1 (AF156100) hemicentin [Homo sapiens]
242	gi13872813	3662	0.0	96	(AJ306906) fibulin-6 [Homo sapiens]
242	gi21707866	1402	e-153	40	(BC034076) CDNA sequence BC034076 [Mus musculus]
243	gi20149229	1097	e-119	100	AF493786_1 (AF493786) koyt binding protein 1 [Homo sapiens]
243	gi20149223	1097	e-119	100	AF493783_1 (AF493783) koyt binding protein 1 [Homo sapiens]
243	gi8052242	1094	e-118	99	(AJ400877) C11orf17 protein [Homo sapiens]
244	gi16553200	1571	e-173	100	(AK057477) unnamed protein product [Homo sapiens]
244	gi15929192	1487	e-163	99	(BC015047) FLJ32915 protein [Homo sapiens]
244	gi23271139	1265	e-138	81	(BC035953) 3010015K02Rik protein [Mus musculus]
245	gi8118086	6532	0.0	80	AF218940_1 (AF218940) formin-2 [Mus musculus]
245	gi8118088	1715	0.0	100	(AF218941) formin 2-like protein [Homo sapiens]
245	gi8118090	1533	e-168	100	(AF218942) formin 2-like protein [Homo sapiens]
246	gi6523831	1800	0.0	100	AF113538_1 (AF113538) retinoid x receptor interacting protein [Homo sapiens]
246	gi12584845	1783	0.0	99	AF284753_1 (AF284753) X2HRIP110 [Homo

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					sapiens]
246	gi21619703	1643	0.0	99	(BC032561) RAP80 protein [Homo sapiens]
248	gi11177164	16715	0.0	81	AF206329_1 (AF206329) polydom protein [Mus musculus]
248	gi14198157	3176	0.0	79	(BC008135) D430029O09Rik protein [Mus musculus]
248	gi12060830	2520	0.0	94	AF308289_1 (AF308289) serologically defined breast cancer antigen NY-BR-38 [Homo sapiens]
249	gi11177164	4047	0.0	83	AF206329_1 (AF206329) polydom protein [Mus musculus]
249	gi182048	329	6e-29	27	(M30640) endothelial leukocyte adhesion molecule 1 [Homo sapiens]
249	gi537524	329	6e-29	27	(M24736) endothelial leukocyte adhesion molecule 1 [Homo sapiens]
250	gi11177164	1975	0.0	80	AF206329_1 (AF206329) polydom protein [Mus musculus]
250	gi7297206	513	1e-50	33	(AE003615) CG9138-PA [Drosophila melanogaster]
250	gi499688	368	9e-34	57	(L33862) fibropellin III [Heliocidaris erythrogramma]
251	gi11177164	12675	0.0	80	AF206329_1 (AF206329) polydom protein [Mus musculus]
251	gi14198157	3176	0.0	79	(BC008135) D430029O09Rik protein [Mus musculus]
251	gi12060830	2520	0.0	94	AF308289_1 (AF308289) serologically defined breast cancer antigen NY-BR-38 [Homo sapiens]
252	gi28422541	2162	0.0	100	(BC047003) PTDSR protein [Homo sapiens]
252	gi11037740	2130	0.0	97	(AF304118) apoptotic cell clearance receptor PtdSerR [Mus musculus]
252	gi34785299	2130	0.0	97	(BC056629) Phosphatidylserine receptor [Mus musculus]
254	gi21615526	2413	0.0	98	(AJ314648) ATP(GTP)-binding protein [Homo sapiens]
254	gi34785807	150	2e-08	24	(BC057535) Unknown (protein for MGC:66453) [Danio rerio]
255	gi15987495	2800	0.0	100	AF378757_1 (AF378757) tumor endothelial marker 7-related precursor [Homo sapiens]
255	gi37182095	2798	0.0	99	(AY358486) ARFP2514 [Homo sapiens]
255	gi34784660	2541	0.0	91	(BC057881) Tumor endothelial marker 7-related precursor [Mus musculus]
256	gi39644911	972	e-104	100	(BC012733) CHMP1.5 protein [Homo sapiens]
256	gi17933108	972	e-104	100	AF306520_1 (AF306520) C18orf2 [Homo sapiens]
256	gi9885435	957	e-102	100	AF281064_1 (AF281064) CHMP1.5 [Homo sapiens]
257	gi32527651	1912	0.0	93	(AY323910) sulfatase modifying factor 1 [Homo sapiens]
257	gi30840149	1912	0.0	93	(AY208752) C-alpha-formyglycine-generating enzyme [Homo sapiens]
257	gi37181290	1718	0.0	91	(AY358092) AAPA3037 [Homo sapiens]
258	gi5764555	908	4e-97	100	AF172331_1 (AF172331) lithostathine [Homo sapiens]
258	gi13529161	908	4e-97	100	(BC005350) Regenerating islet-derived 1 alpha, precursor [Homo sapiens]
258	gi190979	908	4e-97	100	(M18963) islet regenerating protein [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
259	gi42415297	986	e-106	100	(AB158319) membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1 [Homo sapiens]
259	gi12655217	986	e-106	100	(BC001467) SIPL protein [Homo sapiens]
259	gi33150590	982	e-105	99	AF087863_1 (AF087863) submergence induced protein 2 [Homo sapiens]
260	gi19880265	1649	0.0	92	(AF363483) metallo phosphoesterase [Homo sapiens]
260	gi19880267	1649	0.0	92	AF363484_1 (AF363484) metallo phosphoesterase [Homo sapiens]
260	gi19880264	1649	0.0	92	(AF363483) metallo phosphoesterase [Homo sapiens]
261	gi15963593	7806	0.0	100	AF414401_1 (AF414401) ADAMTS13 [Homo sapiens]
261	gi16117338	7806	0.0	100	(AB069698) von Willebrand factor-cleaving protease [Homo sapiens]
261	gi16306598	7802	0.0	99	(AY055376) von Willebrand factor-cleaving protease precursor [Homo sapiens]
262	gi20380757	1565	e-173	100	(BC027867) SLAMF7 protein [Homo sapiens]
262	gi7161175	1410	e-155	100	(AJ271869) 19A24 protein [Homo sapiens]
262	gi14517606	1349	e-148	100	(AB027233) membrane protein FOAP-12 [Homo sapiens]
263	gi10197717	3426	0.0	99	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
263	gi1235698	3180	0.0	97	(L42621) Ly-9 gene product [Homo sapiens]
263	gi10141011	1798	0.0	55	(AF246701) leukocyte cell-surface molecule [Mus musculus]
264	gi9588414	216	4e-17	100	(AL121985) bA404F10.5 (lymphocyte antigen 9) [Homo sapiens]
264	gi40039550	216	4e-17	100	(AX884413) unnamed protein product [Homo sapiens]
264	gi10197717	216	4e-17	100	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
265	gi10197717	3340	0.0	97	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
265	gi1235698	3216	0.0	99	(L42621) Ly-9 gene product [Homo sapiens]
265	gi10141011	1735	0.0	54	(AF246701) leukocyte cell-surface molecule [Mus musculus]
266	gi10197717	3274	0.0	96	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
266	gi1235698	3028	0.0	93	(L42621) Ly-9 gene product [Homo sapiens]
266	gi10141011	1690	0.0	53	(AF246701) leukocyte cell-surface molecule [Mus musculus]
267	gi10197717	3216	0.0	99	AF244129_1 (AF244129) cell-surface molecule Ly-9 [Homo sapiens]
267	gi1235698	3135	0.0	97	(L42621) Ly-9 gene product [Homo sapiens]
267	gi10141011	1706	0.0	55	(AF246701) leukocyte cell-surface molecule [Mus musculus]
268	gi27469556	246	1e-20	42	(BC042054) Putative neuronal cell adhesion molecule [Homo sapiens]
268	gi31418555	234	2e-19	42	(BC053057) Punc protein [Mus musculus]
268	gi3068592	234	2e-19	42	(AF026465) punc [Mus musculus]
269	gi13278924	748	1e-78	98	(BC004217) Neural proliferation, differentiation and control, 1 [Homo sapiens]
269	gi18028281	748	1e-78	98	AF327349_1 (AF327349) NPDC-1 protein [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
269	gi8515886	748	1e-78	98	AF272357_1 (AF272357) NPDC1-like protein [Homo sapiens]
270	gi15929766	1815	0.0	81	(BC015304) Ahcy protein [Mus musculus]
270	gi30584089	1814	0.0	81	(BT007625) Homo sapiens S-adenosylhomocysteine hydrolase [synthetic construct]
270	gi178279	1814	0.0	81	(M61832) S-adenosylhomocysteine hydrolase [Homo sapiens]
271	gi15559823	2253	0.0	89	(BC014258) IGHG1 protein [Homo sapiens]
271	gi17939658	2145	0.0	86	(BC019337) IGHG1 protein [Homo sapiens]
271	gi19684012	2130	0.0	86	(BC026038) IGHG1 protein [Homo sapiens]
272	gi15929988	497	9e-50	100	(BC015423) Family with sequence similarity 14, member B [Homo sapiens]
272	gi21618549	303	3e-27	70	(BC032626) TLH29 protein precursor [Homo sapiens]
272	gi11493982	303	3e-27	70	AF208232_1 (AF208232) TLH29 protein precursor [Homo sapiens]
273	gi9931976	2013	0.0	98	(U29195) neuronal pentraxin II [Homo sapiens]
273	gi881934	2013	0.0	98	(U26662) neuronal pentraxin II [Homo sapiens]
273	gi37574029	1998	0.0	98	(BC048275) Neuronal pentraxin II [Homo sapiens]
274	gi16877407	271	2e-23	66	(BC016950) MGC22679 protein [Homo sapiens]
274	gi28802429	269	3e-23	55	(AX647813) unnamed protein product [Homo sapiens]
274	gi28801684	262	2e-22	52	(AX647581) unnamed protein product [Homo sapiens]
275	gi14280020	3380	0.0	49	(AF312825) collagen type XX alpha 1 precursor [Gallus gallus]
275	gi20988506	2686	0.0	70	(BC030415) 1700051I12Rik protein [Mus musculus]
275	gi288873	1294	e-140	36	(X70793) collagen XIV [Gallus gallus]
276	gi14280020	3652	0.0	52	(AF312825) collagen type XX alpha 1 precursor [Gallus gallus]
276	gi20988506	3056	0.0	76	(BC030415) 1700051I12Rik protein [Mus musculus]
276	gi288873	1294	e-140	36	(X70793) collagen XIV [Gallus gallus]
277	gi14280020	3465	0.0	50	(AF312825) collagen type XX alpha 1 precursor [Gallus gallus]
277	gi20988506	2841	0.0	72	(BC030415) 1700051I12Rik protein [Mus musculus]
277	gi288873	1294	e-140	36	(X70793) collagen XIV [Gallus gallus]
278	gi29126824	915	6e-97	34	(BC047979) MGC53743 protein [Xenopus laevis]
278	gi2258274	876	2e-92	42	(U79775) NNP-1/Nop52 [Homo sapiens]
278	gi7768761	876	2e-92	42	(AP001752) NNP-1/Nop52 (NNP-1), novel nuclear protein 1 [Homo sapiens]
279	gi21901937	2911	0.0	100	(AJ487961) LGI1-like protein 4 [Homo sapiens]
279	gi21359658	2911	0.0	100	(AF467956) LGI3 [Homo sapiens]
279	gi20975686	2911	0.0	100	(AJ487518) leucine-rich glioma inactivated protein 3 [Homo sapiens]
280	gi41396347	141	2e-07	30	(AE017233) FtsQ [Mycobacterium avium subsp. paratuberculosis str. k10]
281	gi15559781	1733	0.0	100	(BC014241) G protein-coupled receptor 146 [Homo sapiens]
281	gi13097087	1273	e-139	74	(BC003323) CDNA sequence BC003323 [Mus musculus]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
281	gi38197529	1092	e-118	60	(BC061674) MGC68817 protein [Xenopus laevis]
282	gi6572272	4157	0.0	100	(AL035681) dJ756G23.1 (novel Leucine Rich Protein) [Homo sapiens]
282	gi29387139	2388	0.0	99	(BC048421) LOC150356 protein [Homo sapiens]
282	gi470672	653	2e-66	41	(U08018) cartilage leucine-rich protein [Bos taurus]
283	gi36603	2198	0.0	99	(Z11773) SRE-ZBP [Homo sapiens]
283	gi15530309	1774	0.0	99	(BC013951) Zinc finger protein 187 [Homo sapiens]
283	gi15530328	1774	0.0	99	(BC013962) Zinc finger protein 187 [Homo sapiens]
284	gi19171178	3590	0.0	79	(AJ315734) metalloprotease disintegrin 16 with thrombospondin type I motif [Homo sapiens]
284	gi21961374	1836	0.0	79	(BC034739) A disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 16 [Mus musculus]
284	gi38649249	1160	e-125	55	(BC063283) ADAMTS18 protein [Homo sapiens]
285	gi8547215	1289	e-141	100	AF205940_1 (AF205940) endomucin [Homo sapiens]
285	gi6252444	1282	e-140	99	(AB034695) endomucin-2 [Homo sapiens]
285	gi21724166	1093	e-118	100	(AY039241) gastric cancer antigen Ga34 [Homo sapiens]
286	gi21320872	2744	0.0	87	(AB041610) Cog8 [Mus musculus]
286	gi7297851	1143	e-123	43	(AE003632) CG6488-PA [Drosophila melanogaster]
286	gi17028369	1139	e-123	100	(BC017492) COG8 protein [Homo sapiens]
287	gi6539606	3918	0.0	99	(AF086645) metastasis suppressor protein [Homo sapiens]
287	gi18848244	3785	0.0	96	(BC024131) Actin monomer-binding protein [Mus musculus]
287	gi28894435	3785	0.0	96	(AY214918) actin monomer-binding protein MIM [Mus musculus]
288	gi18378673	446	8e-44	100	AF462605_1 (AF462605) PATE [Homo sapiens]
288	gi12406754	446	8e-44	100	(AX061647) unnamed protein product [Homo sapiens]
289	gi18378673	608	1e-62	90	AF462605_1 (AF462605) PATE [Homo sapiens]
289	gi12406754	607	2e-62	89	(AX061647) unnamed protein product [Homo sapiens]
290	gi18378673	692	2e-72	100	AF462605_1 (AF462605) PATE [Homo sapiens]
290	gi12406754	691	3e-72	99	(AX061647) unnamed protein product [Homo sapiens]
291	gi28436814	1001	e-107	87	(BC047081) LOC201191 protein [Homo sapiens]
291	gi29437166	923	1e-98	81	(BC049954) CDNA sequence BC034054 [Mus musculus]
291	gi21707603	923	1e-98	81	(BC034054) CDNA sequence BC034054 [Mus musculus]
292	gi22316603	6091	0.0	99	(AX481763) unnamed protein product [Homo sapiens]
292	gi7715417	5114	0.0	85	AF236061_1 (AF236061) RING-finger binding protein [Oryctolagus cuniculus]
292	gi6457274	3340	0.0	56	AF156551_1 (AF156551) putative E1-E2 ATPase [Mus musculus]
293	gi18496663	2676	0.0	100	(AF465771) copine-like protein isoform B [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
293	gi18496661	2676	0.0	100	(AF465770) copine-like protein isoform A [Homo sapiens]
293	gi15680118	2676	0.0	100	(BC014396) Copine IV [Homo sapiens]
294	gi3309151	11773	0.0	99	(AF055136) alpha-tectorin [Homo sapiens]
294	gi1915909	11411	0.0	95	(X99805) alpha tectorin [Mus musculus]
294	gi4049439	8659	0.0	73	(AJ012287) alpha tectorin [Gallus gallus]
295	gi18676472	7210	0.0	99	(AK074062) FLJ00133 protein [Homo sapiens]
295	gi37605781	6106	0.0	83	(AJ584850) secreted nidogen domain protein precursor [Mus musculus]
295	gi29568116	4677	0.0	85	(AY169783) secreted protein SST3 [Mus musculus]
296	gi34528596	593	5e-61	79	(AK123126) unnamed protein product [Homo sapiens]
296	gi23172107	139	2e-08	36	(AE003745) CG33108-PA [Drosophila melanogaster]
296	gi3878329	120	4e-06	32	(Z81097) Hypothetical protein K07A1.3 [Caenorhabditis elegans]
297	gi12832380	1782	0.0	89	(AK002414) unnamed protein product [Mus musculus]
297	gi5441942	1723	0.0	100	AC004997_5 (AC004997) supported by mouse EST AA538043 (NID:g2284036) [Homo sapiens]
297	gi24636593	204	7e-15	28	(AB095109) CiGl [Ciona intestinalis]
298	gi20086516	490	1e-48	100	AF245303_1 (AF245303) prominin-2 variant A [Homo sapiens]
298	gi20086518	490	1e-48	100	AF245304_1 (AF245304) prominin-2 variant B [Homo sapiens]
298	gi37181879	490	1e-48	100	(AY358377) PROM2 [Homo sapiens]
299	gi20086516	3442	0.0	99	AF245303_1 (AF245303) prominin-2 variant A [Homo sapiens]
299	gi20086518	3442	0.0	99	AF245304_1 (AF245304) prominin-2 variant B [Homo sapiens]
299	gi37181879	3442	0.0	99	(AY358377) PROM2 [Homo sapiens]
300	gi20086516	1063	e-115	99	AF245303_1 (AF245303) prominin-2 variant A [Homo sapiens]
300	gi20086518	1063	e-115	99	AF245304_1 (AF245304) prominin-2 variant B [Homo sapiens]
300	gi37181879	1063	e-115	99	(AY358377) PROM2 [Homo sapiens]
301	gi14714659	386	8e-37	100	(BC010469) PEA15 protein [Homo sapiens]
301	gi598187	310	5e-28	82	(L37385) unknown [Homo sapiens]
301	gi473910	141	2e-08	90	(L31958) mammary transforming protein [Mus musculus]
302	gi13195441	896	2e-95	82	AF327440_1 (AF327440) BTE-binding protein 4 [Homo sapiens]
302	gi14549656	731	3e-76	71	AF283891_1 (AF283891) dopamine receptor regulating factor [Mus musculus]
302	gi19919730	528	1e-52	46	AF490374_1 (AF490374) BTEB5 [Homo sapiens]
303	gi29468510	604	3e-62	100	(AY169281) putative fibrinogen-like protein [Homo sapiens]
303	gi37182772	604	3e-62	100	(AY358827) ANGPTL5 [Homo sapiens]
303	gi29351676	604	3e-62	100	(BC049170) Angiopoietin-like 5 [Homo sapiens]
304	gi14164615	2143	0.0	100	AF310234_1 (AF310234) sialic acid binding immunoglobulin-like lectin 8 [Homo sapiens]
304	gi9837433	1320	e-144	96	AF287892_1 (AF287892) sialic acid binding immunoglobulin-like lectin 8 long splice variant

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					[Homo sapiens]
304	gi6289055	1295	e-141	69	AF193441_1 (AF193441) Siglec-7 [Homo sapiens]
305	gi11231111	437	8e-43	74	(AB051124) hypothetical protein [Macaca fascicularis]
306	gi556651	1634	e-180	88	(X78342) PISSLRE [Homo sapiens]
306	gi4490795	1634	e-180	88	(AJ010341) cyclin-dependent kinase [Homo sapiens]
306	gi8521453	1289	e-140	86	(L33264) CDC2-related protein kinase [Homo sapiens]
307	gi7363342	1819	0.0	100	AF193507_1 (AF193507) chemokine receptor [Homo sapiens]
307	gi7328552	1819	0.0	100	AF110640_1 (AF110640) orphan seven-transmembrane receptor [Homo sapiens]
307	gi7274392	1819	0.0	100	(AF233281) CC chemokine receptor [Homo sapiens]
308	gi24817412	877	1e-93	100	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
308	gi40978142	591	2e-60	100	(AX970611) unnamed protein product [Homo sapiens]
308	gi40981860	344	9e-32	100	(AX972470) unnamed protein product [Homo sapiens]
309	gi24817412	853	2e-90	99	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
309	gi40978142	567	2e-57	99	(AX970611) unnamed protein product [Homo sapiens]
309	gi40981860	320	1e-28	98	(AX972470) unnamed protein product [Homo sapiens]
310	gi24817412	264	1e-22	88	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
311	gi24817412	853	1e-90	99	(AF518873) type II transmembrane protein DCAL1 [Homo sapiens]
311	gi40978142	567	2e-57	99	(AX970611) unnamed protein product [Homo sapiens]
311	gi40981860	320	8e-29	98	(AX972470) unnamed protein product [Homo sapiens]
312	gi17940758	3771	0.0	99	AF451977_1 (AF451977) cask-interacting protein 1 [Homo sapiens]
312	gi17940754	3335	0.0	88	AF451975_1 (AF451975) cask-interacting protein 1 [Rattus norvegicus]
312	gi38511409	3312	0.0	88	(BC060720) C630036E02Rik protein [Mus musculus]
313	gi6273399	4573	0.0	59	AF200348_1 (AF200348) melanoma-associated antigen MG50 [Homo sapiens]
313	gi1504040	4573	0.0	59	(D86983) similar to D.melanogaster peroxidasin(U11052) [Homo sapiens]
313	gi7292259	2604	0.0	38	(AE003475) CG12002-PA [Drosophila melanogaster]
314	gi6562060	5211	0.0	98	(AL035659) dJ979N1.1 (dJ979N1.1) [Homo sapiens]
314	gi6176338	4027	0.0	99	AF188530_1 (AF188530) ubiquitous tetratricopeptide containing protein RoXaN [Homo sapiens]
314	gi34783369	3435	0.0	100	(BC024313) RoXaN protein [Homo sapiens]
315	gi15079904	1843	0.0	88	(BC011746) Torsin family 3, member A [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
315	gi14043167	1843	0.0	88	(BC007571) Torsin family 3, member A [Homo sapiens]
315	gi12654511	1843	0.0	88	(BC001085) Torsin family 3, member A [Homo sapiens]
316	gi9716665	2901	0.0	100	(AF282874) nectin 3; PRR3 [Homo sapiens]
316	gi21444080	2901	0.0	100	(AX411429) unnamed protein product [Homo sapiens]
316	gi7546797	2721	0.0	92	AF195833_1 (AF195833) cell adhesion molecule nectin-3 alpha [Mus musculus]
317	gi6289071	1258	e-137	100	AF196972_4 (AF196972) phenylalkylamine binding protein [Homo sapiens]
317	gi6289074	1258	e-137	100	AF196969_1 (AF196969) phenylalkylamine binding protein [Homo sapiens]
317	gi780263	1258	e-137	100	(Z37986) phenylalkylamine binding protein [Homo sapiens]
318	gi37182454	388	4e-37	100	(AY358666) CSRP2BP [Homo sapiens]
318	gi7296222	153	8e-10	50	(AE003590) CG11562-PA [Drosophila melanogaster]
318	gi21429160	153	8e-10	50	(AY119645) RE44650p [Drosophila melanogaster]
319	gi19171211	3367	0.0	100	(AJ421515) CRTAC1-B protein [Homo sapiens]
319	gi10178883	3179	0.0	100	(AJ279016) chondrocyte expressed protein 68 kDa [Homo sapiens]
319	gi9368807	3179	0.0	100	(AJ276171) ASPIC [Homo sapiens]
320	gi30583367	984	e-105	68	(BT007264) interferon regulatory factor 2 [Homo sapiens]
320	gi16041826	984	e-105	68	(BC015803) Interferon regulatory factor 2 [Homo sapiens]
320	gi19387294	960	e-103	65	AF480857_1 (AF480857) interferon regulatory factor 2 [Sigmodon hispidus]
321	gi10444285	1649	0.0	100	(AF290204) blood group carrier molecule DOK1 [Homo sapiens]
321	gi20385811	1649	0.0	100	(AF382213) Dombrock blood group carrier molecule [Homo sapiens]
321	gi20385818	1644	0.0	99	(AF382216) Dombrock blood group carrier molecule [Homo sapiens]
322	gi18535616	5262	0.0	90	(AY074490) EEG1L [Homo sapiens]
322	gi15077418	1385	e-151	100	AF326778_1 (AF326778) gastric cancer multidrug resistance-associated protein [Homo sapiens]
322	gi18535618	1371	e-149	100	(AY074491) EEG1S [Homo sapiens]
323	gi33187657	630	3e-65	100	AF451994_1 (AF451994) ankyrin repeat-containing SOCS box protein 7 [Homo sapiens]
323	gi38614409	621	3e-64	98	(BC062948) AI449039 protein [Mus musculus]
323	gi15420873	615	2e-63	97	AF398968_1 (AF398968) ankyrin repeat-containing SOCS box protein 7 [Mus musculus]
324	gi3746652	964	e-103	100	(AF070523) JWA protein [Homo sapiens]
324	gi6563260	964	e-103	100	AF125530_1 (AF125530) jmx protein [Homo sapiens]
324	gi31455557	964	e-103	100	(AB097051) putative MAPK activating protein [Homo sapiens]
325	gi15779083	1138	e-123	91	(BC014609) IMAGE:4215339 protein [Homo sapiens]
325	gi3342737	983	e-105	88	(AC005328) R26660_2, partial CDS [Homo sapiens]
325	gi37182012	667	7e-69	97	(AY358444) ALLL831 [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
326	gi6180011	1074	e-116	100	AF191338_1 (AF191338) anaphase-promoting complex subunit 4 [Homo sapiens]
326	gi37590799	1067	e-115	99	(BC059383) Anaphase-promoting complex subunit 4 [Homo sapiens]
326	gi19353519	921	2e-98	85	(BC024870) Anaphase-promoting complex subunit 4 [Mus musculus]
327	gi30842594	2218	0.0	96	(AJ318051) putative sulfhydryl oxidase precursor [Homo sapiens]
327	gi34192895	2201	0.0	100	(BC047604) QSCN6L1 protein [Homo sapiens]
327	gi22658418	1999	0.0	83	(BC030934) Quiescin Q6-like 1 [Mus musculus]
328	gi12804553	1592	e-176	100	(BC001689) Carnitine/acylcarnitine translocase [Homo sapiens]
328	gi2765075	1592	e-176	100	(Y10319) carnitine carrier [Homo sapiens]
328	gi5851675	1582	e-175	99	(Y17775) carnitine/acylcarnitine translocase [Homo sapiens]
329	gi38522	1305	e-143	92	(Z21507) human elongation factor-1-delta [Homo sapiens]
329	gi30583323	1302	e-142	92	(BT007242) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
329	gi30584927	1302	e-142	92	(BT008044) Homo sapiens eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [synthetic construct]
330	gi33341656	917	8e-98	73	AF370363_1 (AF370363) FP1047 [Homo sapiens]
330	gi30583323	860	3e-91	84	(BT007242) eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [Homo sapiens]
330	gi30584927	860	3e-91	84	(BT008044) Homo sapiens eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) [synthetic construct]
331	gi20070760	1068	e-115	100	(BC026238) Orosomucoid 1 precursor [Homo sapiens]
331	gi757907	1064	e-115	99	(X02544) alpha1-acid glycoprotein [Homo sapiens]
331	gi178257	1064	e-115	99	(M13692) alpha-1 acid glycoprotein precursor [Homo sapiens]
332	gi17061809	593	6e-61	100	(AY040090) C21orf15 protein [Homo sapiens]
333	gi50619	565	1e-57	100	(X01756) cytochrome c [Mus musculus]
333	gi203723	565	1e-57	100	(M20622) somatic cytochrome c [Rattus norvegicus]
333	gi203699	565	1e-57	100	(K00750) cytochrome c [Rattus norvegicus]
334	gi37181654	2351	0.0	100	(AY358267) PUMPCn [Homo sapiens]
334	gi15418728	2340	0.0	99	(AY008443) six transmembrane prostate protein v1 [Homo sapiens]
334	gi15418732	2290	0.0	99	(AY008445) STAMP1 [Homo sapiens]
335	gi15080288	138	5e-08	100	(BC011906) NIFU protein [Homo sapiens]
335	gi11545707	138	5e-08	100	(AY009128) ISCU2 [Homo sapiens]
335	gi29476869	125	2e-06	93	(BC048409) Nitrogen fixation cluster-like [Mus musculus]
336	gi17224904	1952	0.0	43	AF317839_1 (AF317839) immunoglobulin superfamily member 9 [Mus musculus]
336	gi25955616	1942	0.0	42	(BC040281) Igsf9 protein [Mus musculus]
336	gi20988778	1910	0.0	42	(BC030141) Immunoglobulin superfamily, member 9 [Homo sapiens]
337	gi34785669	1873	0.0	88	(BC057168) 9130020G22Rik protein [Mus

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					musculus]
337	gi28839734	1355	e-148	66	(BC047987) Dj462o23.2-prov protein [Xenopus laevis]
337	gi12654843	1072	e-115	100	(BC001265) DJ462O23.2 protein [Homo sapiens]
338	gi17861384	5677	0.0	100	(AY061759) nesprin-2 gamma [Homo sapiens]
338	gi17016967	5677	0.0	100	AF435011_1 (AF435011) NUANCE [Homo sapiens]
338	gi24417711	5677	0.0	100	(AF495911) nesprin-2 [Homo sapiens]
339	gi32693722	2239	0.0	97	(AX776003) unnamed protein product [Homo sapiens]
339	gi14248997	2239	0.0	97	AF376725_1 (AF376725) lung seven transmembrane receptor 1 [Homo sapiens]
339	gi10047325	2237	0.0	99	(AB046844) KIAA1624 protein [Homo sapiens]
340	gi30354285	2105	0.0	100	(BC051858) Adiponectin receptor 2 [Homo sapiens]
340	gi38018645	2105	0.0	100	(AY424280) progestin and adipoQ receptor family member II [Homo sapiens]
340	gi39795724	1958	0.0	92	(BC064109) Adiponectin receptor 2 [Mus musculus]
341	gi535017	3422	0.0	86	(X76637) tMDC I [Macaca fascicularis]
341	gi1542939	2087	0.0	54	(Y07903) transmembrane protein tMDC I [Rattus norvegicus]
341	gi1666651	2074	0.0	54	(X64227) Cyritestin [Mus musculus]
342	gi212452	182	5e-12	20	(M93676) nonmuscle myosin heavy chain [Gallus gallus]
342	gi212450	182	5e-12	20	(M93676) nonmuscle myosin heavy chain [Gallus gallus]
342	gi212451	182	5e-12	20	(M93676) nonmuscle myosin heavy chain [Gallus gallus]
343	gi22652113	1065	e-115	98	AF406780_1 (AF406780) alpha 1 type XXII collagen [Homo sapiens]
343	gi27469566	1065	e-115	98	(BC042075) COL22A1 protein [Homo sapiens]
343	gi211499	431	1e-41	43	(K01702) HMW/LMW collagen subunit precursor [Gallus gallus]
344	gi825686	4685	0.0	92	(X69301) mast/stem cell growth factor receptor [Homo sapiens]
344	gi1817733	4685	0.0	92	(U63834) KIT protein [Homo sapiens]
344	gi1817734	4647	0.0	92	(U63834) KIT protein [Homo sapiens]
345	gi337934	1376	e-151	96	(M59964) stem cell factor [Homo sapiens]
345	gi15217067	1376	e-151	96	AF400436_1 (AF400436) stem cell factor isoform 1 [Homo sapiens]
345	gi1827477	1195	e-130	84	(D50833) stem cell factor [Felis catus]
346	gi28436366	3508	0.0	99	(AY154461) NALP6 [Homo sapiens]
346	gi19387136	3508	0.0	99	AF479748_1 (AF479748) PYRIN-containing APAF1-like protein 5 [Homo sapiens]
346	gi202806	1566	e-172	67	(M85183) vasopressin receptor [Rattus norvegicus]
347	gi28436366	4563	0.0	99	(AY154461) NALP6 [Homo sapiens]
347	gi19387136	4563	0.0	99	AF479748_1 (AF479748) PYRIN-containing APAF1-like protein 5 [Homo sapiens]
347	gi202806	1566	e-172	67	(M85183) vasopressin receptor [Rattus norvegicus]
348	gi37181742	2425	0.0	100	(AY358311) NL7 [Homo sapiens]
348	gi21432054	2422	0.0	99	(BC032953) Fibrinogen C domain containing 1 [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
348	gi38148657	2216	0.0	90	(BC060634) AI448887 protein [Mus musculus]
349	gi37181742	2450	0.0	100	(AY358311) NL7 [Homo sapiens]
349	gi21432054	2447	0.0	99	(BC032953) Fibrinogen C domain containing 1 [Homo sapiens]
349	gi38148657	2241	0.0	90	(BC060634) AI448887 protein [Mus musculus]
350	gi21667214	2286	0.0	100	AF465767_1 (AF465767) bactericidal/permeability-increasing protein-like 3 [Homo sapiens]
350	gi28856146	1584	e-175	71	(BC048083) Bactericidal/permeability-increasing protein-like 3 precursor [Mus musculus]
350	gi57732	573	1e-57	33	(X60660) potential ligand-binding protein [Rattus rattus]
351	gi13183327	2363	0.0	100	AF274714_1 (AF274714) oxysterol-binding protein-related protein [Homo sapiens]
351	gi39794217	2363	0.0	100	(BC063420) Oxysterol-binding protein-like 1A, isoform A [Homo sapiens]
351	gi17529999	2358	0.0	99	AF392450_1 (AF392450) oxysterol-binding protein-like protein OSBPL1B [Homo sapiens]
352	gi20521035	14493	0.0	100	(AB007859) KIAA0399 protein [Homo sapiens]
352	gi34534413	4574	0.0	99	(AK127482) unnamed protein product [Homo sapiens]
352	gi22766839	1065	e-113	95	(BC037463) C130099L13Rik protein [Mus musculus]
353	gi18381163	1462	e-161	94	(BC022187) C1q and tumor necrosis factor related protein 7 [Homo sapiens]
353	gi18645144	1462	e-161	94	(BC024015) C1q and tumor necrosis factor related protein 7 [Homo sapiens]
353	gi13274524	1462	e-161	94	AF329839_1 (AF329839) complement-c1q tumor necrosis factor-related protein [Homo sapiens]
354	gi21622544	695	9e-73	100	(AJ315533) LY6G6C protein [Homo sapiens]
354	gi5304878	695	9e-73	100	(AJ012008) Ly6-C protein [Homo sapiens]
354	gi4337100	695	9e-73	100	AAD18076 (AF129756) G6c [Homo sapiens]
355	gi10198115	2760	0.0	100	AF279890_1 (AF279890) 2P domain potassium channel TREK2 [Homo sapiens]
355	gi19716290	2690	0.0	99	AF385399_1 (AF385399) potassium channel TREK2 splice variant b [Homo sapiens]
355	gi19716292	2690	0.0	99	AF385400_1 (AF385400) potassium channel TREK2 splice variant c [Homo sapiens]
356	gi19716292	2788	0.0	99	AF385400_1 (AF385400) potassium channel TREK2 splice variant c [Homo sapiens]
356	gi10198115	2697	0.0	100	AF279890_1 (AF279890) 2P domain potassium channel TREK2 [Homo sapiens]
356	gi19716290	2690	0.0	99	AF385399_1 (AF385399) potassium channel TREK2 splice variant b [Homo sapiens]
357	gi37590709	2864	0.0	40	(BC059294) MGC68875 protein [Xenopus laevis]
357	gi177870	2767	0.0	40	(M11313) alpha-2-macroglobulin precursor [Homo sapiens]
357	gi25303946	2767	0.0	40	(BC040071) Alpha 2 macroglobulin precursor [Homo sapiens]
358	gi18138034	2294	0.0	99	(Y19199) paired box protein [Mus musculus]
358	gi1405744	2294	0.0	99	(X63963) Pax-6 (paired box containing gene) [Mus musculus]
358	gi15277449	2294	0.0	99	(BC011272) Paired box gene 6 [Mus musculus]
359	gi37182003	1226	e-133	90	(AY358439) RGNL596 [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
359	gi12652661	1226	e-133	90	(BC000078) Collectin sub-family member 11 [Homo sapiens]
369	gi31455215	1055	e-114	95	(BC009951) Collectin sub-family member 11, isoform b [Homo sapiens]
360	gi37181396	1817	0.0	100	(AY358145) RIWW6503 [Homo sapiens]
360	gi18496364	728	1e-75	46	(AB067770) otolin-1 [Oncorhynchus keta]
360	gi18676606	614	2e-62	41	(AK074129) FLJ00201 protein [Homo sapiens]
361	gi3228237	791	1e-83	69	(AJ006692) ultra high sulfer keratin [Homo sapiens]
361	gi32472	783	9e-83	76	(X63755) high-sulpher keratin [Homo sapiens]
361	gi34223444	782	1e-82	68	(AY360461) UHS KERB-like protein [Homo sapiens]
362	gi3228237	872	6e-93	73	(AJ006692) ultra high sulfer keratin [Homo sapiens]
362	gi200962	823	3e-87	66	(M37759) serine 1 ultra high sulfur protein [Mus musculus]
362	gi34223444	808	2e-85	69	(AY360461) UHS KERB-like protein [Homo sapiens]
363	gi37182231	1832	0.0	96	(AY358554) RPGT208 [Homo sapiens]
363	gi19263589	1802	0.0	96	(BC025407) Layilin [Homo sapiens]
363	gi3790610	1551	e-171	83	(AF093673) layilin [Cricetulus griseus]
365	gi15079904	2154	0.0	100	(BC011746) Torsin family 3, member A [Homo sapiens]
365	gi14043167	2154	0.0	100	(BC007571) Torsin family 3, member A [Homo sapiens]
365	gi12654511	2154	0.0	100	(BC001085) Torsin family 3, member A [Homo sapiens]
366	gi15079904	1843	0.0	88	(BC011746) Torsin family 3, member A [Homo sapiens]
366	gi14043167	1843	0.0	88	(BC007571) Torsin family 3, member A [Homo sapiens]
366	gi12654511	1843	0.0	88	(BC001085) Torsin family 3, member A [Homo sapiens]
368	gi10435784	1011	e-109	100	(AK023755) unnamed protein product [Homo sapiens]
368	gi37181450	1005	e-108	99	(AY358171) APAF6268 [Homo sapiens]
368	gi27451951	1005	e-108	99	(AF534824) TREM-like transcript 2 [Homo sapiens]
369	gi10566471	1375	e-151	99	(AB044560) Gliacolin [Mus musculus]
369	gi14278927	1375	e-151	99	(AB045983) gliacolin [Mus musculus]
369	gi19353133	1375	e-151	99	(BC024634) C1q-like [Mus musculus]
370	gi24371079	1547	e-171	100	(AB046109) CREG2 [Homo sapiens]
370	gi28704036	1539	e-170	99	(BC047514) Cellular repressor of E1A-stimulated genes 2 [Homo sapiens]
370	gi34783235	1539	e-170	99	(BC032949) Cellular repressor of E1A-stimulated genes 2 [Homo sapiens]
371	gi37182207	1207	e-131	99	(AY358542) LAIR hlog [Homo sapiens]
371	gi32396010	179	3e-12	33	(AY247821) immunoglobulin A Fc receptor [Bos taurus]
371	gi6563042	179	3e-12	24	AF109683_1 (AF109683) leukocyte-associated Ig-like receptor 1b [Homo sapiens]
372	gi6563300	260	3e-22	100	AF201951_1 (AF201951) high affinity immunoglobulin epsilon receptor beta subunit [Homo sapiens]
372	gi11559250	260	3e-22	100	(AB026043) MS4A7 [Homo sapiens]
372	gi13655467	260	3e-22	100	AF237916_1 (AF237916) MS4A7 protein

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					[Homo sapiens]
373	gi6690252	236	2e-19	84	AF090944_1 (AF090944) PRO0663 [Homo sapiens]
373	gi34533315	232	5e-19	84	(AK126724) unnamed protein product [Homo sapiens]
373	gi17391109	229	1e-18	82	(BC018471) NFS1 nitrogen fixation 1, isoform b precursor [Homo sapiens]
374	gi31753147	3309	0.0	100	(BC053878) Zeta-chain (TCR) associated protein kinase 70kDa [Homo sapiens]
374	gi20987557	3102	0.0	93	(BC029727) Zeta-chain (TCR) associated protein kinase [Mus musculus]
374	gi1684833	3087	0.0	93	(U77667) tyrosine kinase [Mus musculus]
375	gi18088175	2780	0.0	100	(BC020514) CocoaCrisp [Homo sapiens]
375	gi13241974	2780	0.0	100	AF329197_1 (AF329197) CocoaCrisp [Homo sapiens]
375	gi12002311	2780	0.0	100	AF142573_1 (AF142573) putative secretory protein precursor [Homo sapiens]
376	gi10437229	1803	0.0	100	(AK024825) unnamed protein product [Homo sapiens]
376	gi22832309	185	1e-12	27	(AE003500) CG15916-PA [Drosophila melanogaster]
376	gi18447566	185	1e-12	27	(AY075537) RH08992p [Drosophila melanogaster]
377	gi20988290	781	1e-82	100	(BC029889) Hypothetical protein MGC35169 [Homo sapiens]
377	gi27899965	751	4e-79	99	(AX588218) unnamed protein product [Homo sapiens]
377	gi29437330	343	9e-32	58	(BC049746) 1700018L24Rik protein [Mus musculus]
378	gi20988290	351	8e-33	98	(BC029889) Hypothetical protein MGC35169 [Homo sapiens]
378	gi27899965	321	2e-29	97	(AX588218) unnamed protein product [Homo sapiens]
378	gi27899963	317	7e-29	95	(AX588217) unnamed protein product [Homo sapiens]
379	gi21594969	472	7e-47	100	(BC031610) Hypothetical protein MGC35295 [Homo sapiens]
380	gi16041675	575	7e-59	100	(BC015704) Joined to JAZF1 [Homo sapiens]
380	gi13278157	550	6e-56	94	(BC003922) D11Ertd530e protein [Mus musculus]
380	gi30046920	550	6e-56	94	(BC051099) D11Ertd530e protein [Mus musculus]
381	gi23958222	1975	0.0	99	(BC023635) Lipoic acid synthetase, isoform 1 precursor [Homo sapiens]
381	gi12805345	1787	0.0	90	(BC002141) Lipoic acid synthetase [Mus musculus]
381	gi14669826	1787	0.0	90	(AB057731) lipoic acid synthase [Mus musculus]
382	gi4529898	734	6e-77	82	(AF134726) NG23 [Homo sapiens]
382	gi3986756	485	5e-48	58	(AF109905) NG23 [Mus musculus]
382	gi16118508	485	5e-48	58	AF397036_9 (AF397036) G7d [Mus musculus]
383	gi11066090	1188	e-129	85	AF195192_1 (AF195192) matrix metalloprotease MMP-27 [Homo sapiens]
383	gi37182623	1185	e-128	85	(AY358752) MMP27 [Homo sapiens]
383	gi12006364	1121	e-121	81	AF281673_1 (AF281673) matrix metalloproteinase-27 [Tupaia belangeri]
384	gi24251209	4600	0.0	100	(AY149237) collagen XXVII proalpha 1 chain

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					precursor; preproprotein [Homo sapiens]
384	gi28204656	4147	0.0	89	(AY167568) collagen type XXVII proalpha 1 chain [Mus musculus]
384	gi28172191	4147	0.0	89	(AL683828) bM340H1.1 (novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein) [Mus musculus]
385	gi15215576	2580	0.0	76	(AY050249) BMP-2 inducible kinase [Mus musculus]
385	gi3970852	1132	e-122	100	(AB015331) HRIHFB2017 [Homo sapiens]
385	gi23271902	783	1e-81	98	(BC036021) BMP-2 inducible kinase, isoform b [Homo sapiens]
387	gi13477175	1539	e-170	100	(BC005049) Clone HQ0477 PRO0477p [Homo sapiens]
387	gi14043517	1539	e-170	100	(BC007744) Clone HQ0477 PRO0477p [Homo sapiens]
387	gi6690225	653	4e-67	99	AF090929_2 (AF090929) PRO0477p [Homo sapiens]
388	gi34531772	359	2e-33	66	(AK125618) unnamed protein product [Homo sapiens]
388	gi34526292	356	4e-33	67	(AK129691) unnamed protein product [Homo sapiens]
388	gi10437569	354	6e-33	70	(AK025116) unnamed protein product [Homo sapiens]
389	gi26354052	435	4e-42	59	(AK088927) unnamed protein product [Mus musculus]
389	gi26329371	435	4e-42	59	(AK033677) unnamed protein product [Mus musculus]
389	gi12843048	343	2e-31	72	(AK008696) unnamed protein product [Mus musculus]
390	gi26354052	436	3e-42	55	(AK088927) unnamed protein product [Mus musculus]
390	gi26329371	435	5e-42	59	(AK033677) unnamed protein product [Mus musculus]
390	gi12843048	343	2e-31	72	(AK008696) unnamed protein product [Mus musculus]
392	gi37573961	1792	0.0	100	(BC051875) Putative purinergic receptor P2Y10 [Homo sapiens]
392	gi2104787	1792	0.0	100	(AF000545) putative purinergic receptor P2Y10 [Homo sapiens]
392	gi30526091	1792	0.0	100	(AY275461) putative purinergic receptor P2Y10 [Homo sapiens]
393	gi37573961	1792	0.0	100	(BC051875) Putative purinergic receptor P2Y10 [Homo sapiens]
393	gi2104787	1792	0.0	100	(AF000545) putative purinergic receptor P2Y10 [Homo sapiens]
393	gi30526091	1792	0.0	100	(AY275461) putative purinergic receptor P2Y10 [Homo sapiens]
394	gi19575509	1440	e-158	100	(AX380599) unnamed protein product [Homo sapiens]
394	gi19575655	1440	e-158	100	(AX380745) unnamed protein product [Homo sapiens]
394	gi37181903	1435	e-158	99	(AY358389) VSAA731 [Homo sapiens]
395	gi13539688	2253	0.0	100	AF242530_1 (AF242530) protein kinase C and casein kinase substrate 3 [Homo sapiens]
395	gi15080241	2253	0.0	100	(BC011889) Protein kinase C and casein kinase substrate in neurons 3 [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
395	gi11127646	2253	0.0	100	AF149825_1 (AF149825) PACSIN3 [Homo sapiens]
396	gi7672784	2557	0.0	99	AF143723_1 (AF143723) heat shock protein HSP60 [Homo sapiens]
396	gi6563208	2554	0.0	99	AF112210_1 (AF112210) heat shock protein hsp70-related protein [Homo sapiens]
396	gi12805195	2370	0.0	90	(BC002056) Heat shock protein 4 [Mus musculus]
397	gi21961634	720	1e-74	36	(BC034671) CEACAM5 protein [Homo sapiens]
397	gi180223	717	3e-74	36	(M29540) carcinoembryonic antigen [Homo sapiens]
397	gi178677	717	3e-74	36	(M17303) carcinoembryonic antigen precursor [Homo sapiens]
398	gi21961634	465	4e-45	32	(BC034671) CEACAM5 protein [Homo sapiens]
398	gi180211	462	9e-45	32	(M59710) carcinoembryonic antigen [Homo sapiens]
398	gi178677	462	9e-45	32	(M17303) carcinoembryonic antigen precursor [Homo sapiens]
399	gi21961634	445	1e-42	34	(BC034671) CEACAM5 protein [Homo sapiens]
399	gi180211	442	3e-42	33	(M59710) carcinoembryonic antigen [Homo sapiens]
399	gi178677	442	3e-42	33	(M17303) carcinoembryonic antigen precursor [Homo sapiens]
400	gi26278978	2199	0.0	54	(AY158688) ADAM4 [Mus musculus]
400	gi965014	1407	e-154	53	(U22058) ADAM 4 protein precursor [Mus musculus]
400	gi1061159	1277	e-139	37	(X87205) testicular Metalloprotease-like, Disintegrin-like, Cysteine-rich protein IVa [Macaca fascicularis]
401	gi26278978	777	2e-81	53	(AY158688) ADAM4 [Mus musculus]
401	gi1061163	498	4e-49	43	(X87207) testicular Metalloprotease-like, Disintegrin-like, Cysteine-rich protein IVc [Macaca fascicularis]
401	gi1061161	496	7e-49	42	(X87206) testicular Metalloprotease-like, Disintegrin-like, Cysteine-rich protein IVb [Macaca fascicularis]
402	gi177829	2151	0.0	99	(K01396) alpha-1-antitrypsin [Homo sapiens]
402	gi11493443	2151	0.0	99	AF130117_27 (AF130068) PRO2209 [Homo sapiens]
402	gi28966	2151	0.0	99	(X01683) alpha 1-antitrypsin [Homo sapiens]
403	gi6467202	3321	0.0	99	(AB021642) gonadotropin inducible transcription repressor-2 [Homo sapiens]
403	gi21595832	2531	0.0	71	(BC032753) Zinc finger protein 443 [Homo sapiens]
403	gi4519270	2531	0.0	71	(AB011414) Kruppel-type zinc finger protein [Homo sapiens]
404	gi12804197	1084	e-117	80	(BC002956) Endopeptidase Clp precursor [Homo sapiens]
404	gi963048	1084	e-117	80	(Z50853) CLPP [Homo sapiens]
404	gi3559935	817	3e-86	66	(AJ005253) ClpP protease [Mus musculus]
405	gi219535	564	2e-57	81	(D90277) nonspecific cross-reacting antigen [Homo sapiens]
405	gi180227	560	7e-57	80	(L00692) carcinoembryonic antigen [Homo sapiens]
405	gi3851200	404	9e-39	60	(AC005955) CGM7_HUMAN [Homo sapiens]
406	gi15214636	1319	e-144	100	(BC012444) Chloride intracellular channel 4

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					[Homo sapiens]
406	gi5052202	1305	e-143	99	AF097330_1 (AF097330) H1 chloride channel; p64H1; CLIC4 [Homo sapiens]
406	gi6606085	1304	e-142	98	AF102578_1 (AF102578) intracellular chloride channel protein [Mus musculus]
408	gi6525071	2611	0.0	97	(AF159548) nuclear FMRP interacting protein 1 [Homo sapiens]
408	gi33525186	1806	0.0	69	(BC056192) Nuclear fragile X mental retardation protein interacting protein [Mus musculus]
408	gi6525073	1806	0.0	69	(AF159549) nuclear FMRP interacting protein 1 [Mus musculus]
409	gi32967229	705	6e-74	100	(AY325115) TAFA2 [Homo sapiens]
409	gi32967241	691	3e-72	96	(AY325121) TAFA2 [Mus musculus]
409	gi32967233	473	5e-47	69	(AY325117) TAFA4 [Homo sapiens]
410	gi14336713	3060	0.0	100	AE006464_13 (AE006464) possible G-protein receptor [Homo sapiens]
410	gi5912459	1110	e-119	100	(Z97653) c380A1.1 (novel protein) [Homo sapiens]
410	gi19528545	1053	e-113	52	(AY089649) RH06780p [Drosophila melanogaster]
411	gi29373914	912	1e-97	100	(AY158895) alpha 1 type XXIII collagen [Homo sapiens]
411	gi29373916	893	2e-95	97	(AY158896) alpha 1 type XXIII collagen [Rattus norvegicus]
411	gi22652221	889	5e-95	96	AF410792_1 (AF410792) alpha 1 type XXIII collagen [Mus musculus]
412	gi25992504	3884	0.0	79	(AF525689) signal peptide-CUB-EGF-like domain containing protein 1 [Homo sapiens]
412	gi10998440	3167	0.0	69	AF276425_1 (AF276425) EGF-related protein SCUBE1 [Mus musculus]
412	gi8052237	2916	0.0	58	(AJ400877) CEGP1 protein [Homo sapiens]
413	gi25992504	3868	0.0	79	(AF525689) signal peptide-CUB-EGF-like domain containing protein 1 [Homo sapiens]
413	gi10998440	3151	0.0	69	AF276425_1 (AF276425) EGF-related protein SCUBE1 [Mus musculus]
413	gi8052237	2898	0.0	58	(AJ400877) CEGP1 protein [Homo sapiens]
414	gi33285263	294	3e-26	77	(AY236503) cytochrome c oxidase subunit VIc [Tarsius syrichta]
414	gi33285281	267	4e-23	69	(AY236512) cytochrome c oxidase subunit VIc [Nycticebus coucang]
414	gi203519	251	3e-21	68	(M27466) cytochrome c oxidase subunit VIc [Rattus norvegicus]
415	gi37181414	1267	e-138	97	(AY358153) AWKS9372 [Homo sapiens]
415	gi61	158	8e-10	28	(X16451) calmodulin-independent adenylate cyclase [Bos taurus]
415	gi28703938	157	1e-09	28	(BC047244) Neural cell adhesion molecule 1 [Homo sapiens]
416	gi8118227	1311	e-143	100	(AF231922) C21orf62 protein [Homo sapiens]
416	gi23342580	983	e-105	91	(AX497196) unnamed protein product [Homo sapiens]
416	gi21432076	641	8e-66	58	(BC032975) 4932438H23Rik protein [Mus musculus]
417	gi34783508	1205	e-130	83	(BC038564) FLJ31052 protein [Homo sapiens]
417	gi19569541	353	6e-32	42	AF485812_1 (AF485812) Fc gamma receptor I [Macaca fascicularis]
417	gi292169	351	1e-31	41	(L03418) Fc gamma receptor I [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
418	gi21205864	1591	e-176	100	AF385435_1 (AF385435) T-cell activation protein phosphatase 2C; TA-PP2C [Homo sapiens]
418	gi34100337	1561	e-172	99	(AY357944) T-cell activation protein phosphatase 2C-like protein [Homo sapiens]
418	gi21464366	758	3e-79	52	(AY121659) RE06653p [Drosophila melanogaster]
419	gi190568	1476	e-162	87	(M94890) pregnancy-specific beta-1 glycoprotein [Homo sapiens]
419	gi609318	1475	e-162	88	(U18469) pregnancy-specific beta 1-glycoprotein 4 precursor [Homo sapiens]
419	gi190647	1470	e-162	85	(M69245) pregnancy-specific beta-1-glycoprotein [Homo sapiens]
420	gi38511474	604	3e-62	97	(BC062570) CDH26 protein [Homo sapiens]
420	gi7981304	575	7e-59	84	(AL109928) dJ551D2.1.2 (Cadherin-like 26, variant 2) [Homo sapiens]
420	gi9622236	272	1e-23	100	AF169690_1 (AF169690) cadherin-like protein VR20 [Homo sapiens]
421	gi29650885	991	e-106	99	(AY245915) high density lipoprotein-binding protein [Homo sapiens]
421	gi39795445	980	e-105	98	(BC063857) High density lipoprotein-binding protein [Homo sapiens]
421	gi24817754	465	1e-45	55	(AB095543) high density lipoprotein binding protein 1 [Mus musculus]
423	gi37181871	1818	0.0	98	(AY358373) LHPE306 [Homo sapiens]
423	gi31322514	1350	e-148	73	(AY223873) mannose receptor precursor-like isoform 6 [Mus musculus]
423	gi31322510	1350	e-148	73	(AY223871) mannose receptor precursor-like isoform 4 [Mus musculus]
424	gi13375149	961	e-103	100	(AL109964) dJ1118M15.2 (Novel protein) [Homo sapiens]
424	gi7259265	314	4e-28	50	(AB030198) contains transmembrane (TM) region [Mus musculus]
424	gi20072584	305	5e-27	40	(BC027127) CDNA sequence BC027127 [Mus musculus]
425	gi28279464	1008	e-108	79	(BC046311) Olfactory receptor 70 [Mus musculus]
425	gi32032894	1007	e-108	79	(AY317362) olfactory receptor GA_x6K02T2N78B-16239704-16240654 [Mus musculus]
425	gi18480302	1007	e-108	79	(AY073502) olfactory receptor MOR262-10 [Mus musculus]
426	gi21622561	1086	e-117	100	(AJ315545) LY6G5B protein [Homo sapiens]
426	gi5701854	794	9e-84	100	(AJ245417) LY6G5b protein [Homo sapiens]
426	gi6137324	789	4e-83	99	AF129756_1 (AF129756) G5b [Homo sapiens]
427	gi38382767	491	5e-49	100	(BC062368) Unknown (protein for MGC:71256) [Homo sapiens]
427	gi12652993	491	5e-49	100	(BC000257) LOC152217 protein [Homo sapiens]
427	gi18204855	340	1e-31	75	(BC021536) Unknown (protein for MGC:35773) [Mus musculus]
428	gi40782699	503	2e-50	100	(AX952340) unnamed protein product [Homo sapiens]
428	gi21432071	307	1e-27	65	(BC032982) Unknown (protein for MGC:41689) [Mus musculus]
429	gi20521047	8738	0.0	99	(AB007883) KIAA0423 [Homo sapiens]
429	gi7296250	223	3e-16	31	(AE003590) CG4648-PA [Drosophila

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					melanogaster]
429	gi21064295	223	3e-16	31	(AY113372) LP02990p [Drosophila melanogaster]
430	gi178991	1213	e-132	98	(M83751) arginine-rich protein [Homo sapiens]
430	gi30585119	952	e-102	100	(BT008140) Homo sapiens arginine-rich, mutated in early stage tumors [synthetic construct]
430	gi30583059	952	e-102	100	(BT007110) arginine-rich, mutated in early stage tumors [Homo sapiens]
431	gi19353157	862	2e-91	91	(BC024945) 9430016H08Rik protein [Mus musculus]
431	gi5020383	223	3e-17	32	(AF153450) juvenile hormone esterase binding protein [Manduca sexta]
431	gi17944240	169	6e-11	25	(AY070543) LD24657p [Drosophila melanogaster]
432	gi28208164	533	6e-54	100	(AB081838) secreted Ly6/uPAR related protein 2 [Homo sapiens]
432	gi37181959	533	6e-54	100	(AY358417) QLGT871 [Homo sapiens]
432	gi37572250	460	2e-45	88	(BC032306) Ly-6 neurotoxin-like protein 1, isoform a [Homo sapiens]
434	gi30314483	3584	0.0	99	(AB094094) DLNB23 [Homo sapiens]
434	gi20521025	3343	0.0	100	(AB006623) No similarities to any reported proteins [Homo sapiens]
434	gi37805313	3304	0.0	90	(BC060156) 1300006O23Rik protein [Mus musculus]
435	gi27763975	2569	0.0	100	(AJ312332) APG4-D protein [Homo sapiens]
435	gi27763977	2181	0.0	86	(AJ312333) APG4-D protein [Mus musculus]
435	gi22658287	2177	0.0	85	(BC030861) APG4-D protein [Mus musculus]
436	gi300091	2009	0.0	87	(S59493) pregnancy-specific beta 1-glycoprotein; PSG [Homo sapiens]
436	gi190649	2009	0.0	87	(M93061) pregnancy-specific beta-1 glycoprotein [Homo sapiens]
436	gi180235	2008	0.0	87	(M37399) carcinoembryonic antigen SG5 [Homo sapiens]
437	gi15214951	1553	e-171	87	(BC012607) Pregnancy specific beta-1-glycoprotein 5 [Homo sapiens]
437	gi190634	1534	e-169	86	(M73713) pregnancy-specific beta-1-glycoprotein 5 [Homo sapiens]
437	gi190638	1532	e-169	86	(M25384) fetal liver non-specific cross-reactive antigen-3 precursor protein [Homo sapiens]
438	gi306801	1899	0.0	86	(M34420) pregnancy-specific beta-1 glycoprotein precursor [Homo sapiens]
438	gi306802	1899	0.0	86	(M23575) pregnancy-specific beta-1 glycoprotein precursor [Homo sapiens]
438	gi180235	1899	0.0	86	(M37399) carcinoembryonic antigen SG5 [Homo sapiens]
439	gi20987759	2432	0.0	100	(BC030262) ADAM-TS related protein 1, isoform 3 [Homo sapiens]
439	gi37181773	2362	0.0	95	(AY358327) ADAMTSL1 [Homo sapiens]
439	gi15099921	2352	0.0	95	AF176313_1 (AF176313) ADAM-TS related protein 1 [Homo sapiens]
440	gi37181773	2917	0.0	99	(AY358327) ADAMTSL1 [Homo sapiens]
440	gi15099921	2907	0.0	99	AF176313_1 (AF176313) ADAM-TS related protein 1 [Homo sapiens]
440	gi13183078	2432	0.0	62	AF237652_1 (AF237652) a disintegrin-like and metalloprotease domain with thrombospondin

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
					type I motifs-like 3 [Homo sapiens]
441	gi37181773	2808	0.0	99	(AY358327) ADAMTSL1 [Homo sapiens]
441	gi15099921	2798	0.0	99	AF176313_1 (AF176313) ADAM-TS related protein 1 [Homo sapiens]
441	gi13183078	2484	0.0	60	AF237652_1 (AF237652) a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 [Homo sapiens]
442	gi1536902	560	4e-57	100	(X99977) ARS [Homo sapiens]
442	gi4218459	400	2e-38	69	(AJ132356) ARS component B precursor [Mus musculus]
442	gi37181989	204	9e-16	42	(AY358432) RGTR430 [Homo sapiens]
443	gi2739294	658	2e-68	100	(Y12642) E48 antigen [Homo sapiens]
443	gi21411513	658	2e-68	100	(BC031330) Lymphocyte antigen 6 complex, locus D [Homo sapiens]
443	gi887454	653	7e-68	99	(X82693) E48 antigen [Homo sapiens]
444	gi2739294	287	2e-25	96	(Y12642) E48 antigen [Homo sapiens]
444	gi21411513	287	2e-25	96	(BC031330) Lymphocyte antigen 6 complex, locus D [Homo sapiens]
444	gi887454	282	9e-25	94	(X82693) E48 antigen [Homo sapiens]
445	gi33086556	999	e-107	97	(AY325189) Ab2-095 [Rattus norvegicus]
445	gi21428872	129	6e-06	25	(AY119501) GH11358p [Drosophila melanogaster]
445	gi21626538	129	6e-06	25	(AE003456) CG11170-PB [Drosophila melanogaster]
446	gi13358942	3017	0.0	99	(AB056426) hypothetical protein [Macaca fascicularis]
446	gi37181749	2665	0.0	100	(AY358315) GFNV803 [Homo sapiens]
446	gi29540625	2665	0.0	100	(AY182028) leucine-rich repeat transmembrane neuronal 3 protein [Homo sapiens]
447	gi2913997	1829	0.0	100	(D86359) CD33L2 [Homo sapiens]
447	gi2913995	1742	0.0	100	(D86358) CD33L1 [Homo sapiens]
447	gi20258598	1742	0.0	100	(AY040542) sialic acid binding immunoglobulin-like lectin 6 [Homo sapiens]
448	gi4755085	7197	0.0	99	(AF017178) pro alpha 1(I) collagen [Homo sapiens]
448	gi1418928	7194	0.0	99	(Z74615) prepro-alpha1(I) collagen [Homo sapiens]
448	gi22328092	7175	0.0	99	(BC036531) Alpha 1 type I collagen preproprotein [Homo sapiens]
449	gi6694394	818	8e-87	100	AF201833_1 (AF201833) FIL1 eta [Homo sapiens]
449	gi19068188	516	9e-52	64	(AY071842) IL-1F8 [Mus musculus]
449	gi7769116	452	2e-44	94	AF200494_1 (AF200494) interleukin-1 homolog 2 [Homo sapiens]
450	gi38423520	278	2e-24	47	(AB073023) transmembrane serine protease-1 [Rattus norvegicus]
450	gi26007900	278	2e-24	59	(BC040348) Distal intestinal serine protease [Mus musculus]
450	gi5921501	278	2e-24	59	(AJ243866) distal intestinal serine protease [Mus musculus]
451	gi26007900	1001	e-107	61	(BC040348) Distal intestinal serine protease [Mus musculus]
451	gi15012124	1001	e-107	61	(BC010970) Distal intestinal serine protease [Mus musculus]
451	gi5921501	991	e-106	61	(AJ243866) distal intestinal serine protease [Mus musculus]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
452	gi29126954	1948	0.0	99	(BC047602) RTTN protein [Homo sapiens]
452	gi34783651	1941	0.0	99	(BC046222) RTTN protein [Homo sapiens]
452	gi23271829	1657	0.0	83	(BC023916) Rttn protein [Mus musculus]
453	gi18676606	3953	0.0	100	(AK074129) FLJ00201 protein [Homo sapiens]
453	gi40675467	3768	0.0	94	(BC065148) Procollagen, type VIII, alpha 2 [Mus musculus]
453	gi177179	3520	0.0	97	(M60832) alpha-2 type VIII collagen [Homo sapiens]
454	gi27696986	150	2e-09	43	(BC043846) Armet protein [Xenopus laevis]
454	gi30585119	148	3e-09	59	(BT008140) Homo sapiens arginine-rich, mutated in early stage tumors [synthetic construct]
454	gi178991	148	3e-09	59	(M83751) arginine-rich protein [Homo sapiens]
455	gi21753515	130	3e-07	55	(AK094450) unnamed protein product [Homo sapiens]
456	gi205250	144	8e-09	44	(M30690) Ly6C antigen [Rattus norvegicus]
456	gi1695690	142	1e-08	42	(D86232) Ly-6C variant [Mus musculus]
456	gi198924	139	3e-08	40	(M74013) Ly-6A.2 [Mus musculus]
457	gi13447753	4277	0.0	100	AF296673_1 (AF296673) toll-like receptor 10 [Homo sapiens]
457	gi37181720	4272	0.0	99	(AY358300) TLR10 [Homo sapiens]
457	gi11385997	1937	0.0	50	AF316985_1 (AF316985) toll-like receptor 1 [Mus musculus]
458	gi18378673	196	1e-14	76	AF462605_1 (AF462605) PATE [Homo sapiens]
459	gi12406754	195	2e-14	73	(AX061647) unnamed protein product [Homo sapiens]
460	gi37181989	665	3e-69	100	(AY358432) RGTR430 [Homo sapiens]
460	gi4218459	219	1e-17	44	(AJ132356) ARS component B precursor [Mus musculus]
460	gi1536902	204	8e-16	42	(X99977) ARS [Homo sapiens]
462	gi535017	3379	0.0	83	(X76637) tMDC I [Macaca fascicularis]
462	gi1542939	2050	0.0	52	(Y07903) transmembrane protein tMDC I [Rattus norvegicus]
462	gi1666651	2031	0.0	52	(X64227) Cyritestin [Mus musculus]
463	gi535017	1517	e-167	83	(X76637) tMDC I [Macaca fascicularis]
463	gi1666651	1032	e-111	57	(X64227) Cyritestin [Mus musculus]
463	gi38511880	1007	e-108	57	(BC060975) A disintegrin and metalloprotease domain 3 (cyritestin) [Mus musculus]
464	gi531478	1487	e-163	76	(X77619) tMDC II [Macaca fascicularis]
464	gi965006	943	e-100	50	(U22060) ADAM 5 protein precursor [Cavia porcellus]
464	gi965016	844	4e-89	44	(U22059) ADAM 5 protein precursor [Mus musculus]
465	gi531478	1208	e-131	82	(X77619) tMDC II [Macaca fascicularis]
465	gi965006	804	2e-84	56	(U22060) ADAM 5 protein precursor [Cavia porcellus]
465	gi965016	678	7e-70	47	(U22059) ADAM 5 protein precursor [Mus musculus]
466	gi338294	589	3e-60	53	(M82968) sperm protein 10 [Homo sapiens]
466	gi15779024	589	3e-60	53	(BC014588) Acrosomal vesicle protein 1, isoform a precursor [Homo sapiens]
466	gi7705047	581	2e-59	53	(S65583) SP-10 [Homo sapiens]
467	gi338292	771	3e-81	66	(M82967) sperm protein 10 [Homo sapiens]
467	gi338294	741	1e-77	61	(M82968) sperm protein 10 [Homo sapiens]
467	gi15779024	741	1e-77	61	(BC014588) Acrosomal vesicle protein 1, isoform a precursor [Homo sapiens]

TABLE 2B

SEQ ID	Hit ID	B_score	P_value	% Identity	Annotation
468	gi338294	865	5e-92	69	(M82968) sperm protein 10 [Homo sapiens]
468	gi15779024	865	5e-92	69	(BC014588) Acrosomal vesicle protein 1, isoform a precursor [Homo sapiens]
468	gi7705047	857	4e-91	68	(S65583) SP-10 [Homo sapiens]
469	gi338294	746	2e-78	62	(M82968) sperm protein 10 [Homo sapiens]
469	gi7705047	746	2e-78	62	(S65583) SP-10 [Homo sapiens]
469	gi15779024	746	2e-78	62	(BC014588) Acrosomal vesicle protein 1, isoform a precursor [Homo sapiens]
470	gi338292	468	2e-46	83	(M82967) sperm protein 10 [Homo sapiens]
470	gi298489	464	6e-46	79	(S56458) SP-10 [Papio hamadryas] [Papio papio]
470	gi338294	459	2e-45	82	(M82968) sperm protein 10 [Homo sapiens]

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
237	UCH-1	1/1	41-72	46.3	4.6e-12	Ubiquitin carboxyl-terminal hydrolases famil
237	UCH-2	1/2	285-319	21.8	2e-05	Ubiquitin carboxyl-terminal hydrolase family
237	UCH-2	2/2	448-481	31.3	1.9e-08	Ubiquitin carboxyl-terminal hydrolase family
238	ig	1/2	31-89	26.6	1e-05	Immunoglobulin domain
238	ig	2/2	126-182	21.1	0.00035	Immunoglobulin domain
241	hormone	1/1	9-215	298.6	1.9e-110	Somatotropin hormone family
242	tsp_1	1/3	16-66	59.7	2.5e-16	Thrombospondin type 1 domain
242	tsp_1	2/3	73-123	41.1	8.7e-11	Thrombospondin type 1 domain
242	tsp_1	3/3	130-180	54.7	7.7e-15	Thrombospondin type 1 domain
242	EGF	4/10	423-457	30.9	4.9e-07	EGF-like domain
242	EGF	5/10	463-502	9.8	0.46	EGF-like domain
242	EGF	6/10	508-540	21.9	0.00017	EGF-like domain
242	EGF	8/10	588-625	23.6	5.6e-05	EGF-like domain
242	EGF	9/10	631-665	37.0	9.5e-09	EGF-like domain
245	FH2	2/2	1140-1544	291.8	8.7e-84	Formin Homology 2 Domain
248	vwa	1/1	83-259	82.6	3.8e-23	von Willebrand factor type A domain
248	sushi	1/35	378-433	33.9	3.3e-07	Sushi domain (SCR repeat)
248	sushi	2/35	438-493	58.3	1.2e-13	Sushi domain (SCR repeat)
248	sushi	3/35	498-559	12.7	0.13	Sushi domain (SCR repeat)
248	HYR	1/2	561-642	65.4	3.3e-17	HYR domain
248	HYR	2/2	643-722	65.3	3.6e-17	HYR domain
248	TNFR_c6	3/5	1018-1042	11.5	0.054	
248	TNFR_c6	5/5	1110-1126	8.5	0.46	
248	EGF	4/13	1197-1228	35.5	2.5e-08	EGF-like domain
248	EGF	5/13	1235-1266	45.0	5e-11	EGF-like domain
248	EGF	6/13	1273-1304	34.9	3.6e-08	EGF-like domain
248	EGF	7/13	1311-1342	35.1	3.2e-08	EGF-like domain
248	EGF	8/13	1349-1380	40.4	1e-09	EGF-like domain
248	EGF	9/13	1387-1418	44.6	6.7e-11	EGF-like domain
248	pentaxin	1/1	1470-1608	80.5	2.7e-22	Pentaxin family
248	sushi	5/35	1631-1685	47.3	9.8e-11	Sushi domain (SCR repeat)
248	sushi	6/35	1690-1743	68.8	1.2e-16	Sushi domain (SCR repeat)
248	EGF	10/13	1749-1783	30.0	8.8e-07	EGF-like domain
248	sushi	7/35	1789-1842	62.9	7e-15	Sushi domain (SCR repeat)
248	sushi	8/35	1847-1900	58.5	1.1e-13	Sushi domain (SCR repeat)
248	sushi	9/35	1905-1958	57.5	2.1e-13	Sushi domain (SCR repeat)

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
248	sushi	10/35	1963-2016	56.3	4.1e-13	Sushi domain (SCR repeat)
248	sushi	11/35	2021-2078	30.6	2.5e-06	Sushi domain (SCR repeat)
248	sushi	12/35	2083-2141	39.4	1.2e-08	Sushi domain (SCR repeat)
248	sushi	13/35	2146-2199	71.9	1.3e-17	Sushi domain (SCR repeat)
248	sushi	14/35	2204-2256	48.3	5.2e-11	Sushi domain (SCR repeat)
248	sushi	15/35	2264-2318	67.3	3.3e-16	Sushi domain (SCR repeat)
248	sushi	16/35	2323-2376	38.9	1.5e-08	Sushi domain (SCR repeat)
248	sushi	17/35	2381-2435	56.2	4.3e-13	Sushi domain (SCR repeat)
248	sushi	18/35	2440-2493	48.6	4.3e-11	Sushi domain (SCR repeat)
248	sushi	19/35	2498-2551	62.1	1.2e-14	Sushi domain (SCR repeat)
248	sushi	20/35	2556-2608	53.8	1.9e-12	Sushi domain (SCR repeat)
248	sushi	22/35	2660-2712	51.8	6.4e-12	Sushi domain (SCR repeat)
248	Parameciu m_SA	5/7	2704-2718	8.5	0.14	Paramecium_SA domain
248	sushi	23/35	2717-2770	44.0	7.2e-10	Sushi domain (SCR repeat)
248	sushi	24/35	2775-2828	58.2	1.4e-13	Sushi domain (SCR repeat)
248	sushi	25/35	2833-2886	60.4	3.4e-14	Sushi domain (SCR repeat)
248	sushi	26/35	2891-2944	51.0	1.1e-11	Sushi domain (SCR repeat)
248	sushi	27/35	2949-3002	54.3	1.4e-12	Sushi domain (SCR repeat)
248	sushi	28/35	3007-3059	38.7	1.8e-08	Sushi domain (SCR repeat)
248	sushi	29/35	3064-3117	48.1	6.2e-11	Sushi domain (SCR repeat)
248	sushi	30/35	3122-3176	47.1	1.1e-10	Sushi domain (SCR repeat)
248	sushi	31/35	3181-3230	31.4	1.5e-06	Sushi domain (SCR repeat)
248	sushi	32/35	3241-3294	53.7	2e-12	Sushi domain (SCR repeat)
248	sushi	33/35	3299-3352	46.6	1.5e-10	Sushi domain (SCR repeat)
248	sushi	34/35	3357-3411	42.1	2.3e-09	Sushi domain (SCR repeat)
248	sushi	35/35	3416-3468	53.3	2.6e-12	Sushi domain (SCR repeat)
248	EGF	11/13	3468-3499	21.6	0.00021	EGF-like domain
248	EGF	12/13	3504-3531	29.8	9.9e-07	EGF-like domain

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
248	EGF	13/13	3536-3563	22.5	0.00012	EGF-like domain
249	vwa	1/1	83-259	82.6	3.8e-23	von Willebrand factor type A domain
249	sushi	1/4	378-433	33.9	3.3e-07	Sushi domain (SCR repeat)
249	sushi	2/4	438-493	58.3	1.2e-13	Sushi domain (SCR repeat)
249	sushi	3/4	498-559	12.7	0.13	Sushi domain (SCR repeat)
249	HYR	1/2	561-642	65.4	3.3e-17	HYR domain
249	HYR	2/2	643-722	65.3	3.6e-17	HYR domain
250	TNFR_c6	2/4	153-177	11.5	0.054	TNFR/NGFR cysteine-rich region
250	TNFR_c6	4/4	245-261	8.5	0.46	TNFR/NGFR cysteine-rich region
250	EGF	1/3	332-363	35.5	2.5e-08	EGF-like domain
250	EGF	2/3	370-401	45.0	5e-11	EGF-like domain
250	EGF	3/3	408-437	27.3	5e-06	EGF-like domain
251	TNFR_c6	2/4	153-177	11.5	0.054	TNFR/NGFR cysteine-rich region
251	TNFR_c6	4/4	245-261	8.5	0.46	TNFR/NGFR cysteine-rich region
251	EGF	1/10	332-363	35.5	2.5e-08	EGF-like domain
251	EGF	2/10	370-401	45.0	5e-11	EGF-like domain
251	EGF	3/10	408-439	34.9	3.6e-08	EGF-like domain
251	EGF	4/10	446-477	35.1	3.2e-08	EGF-like domain
251	EGF	5/10	484-515	40.4	1e-09	EGF-like domain
251	EGF	6/10	522-553	44.6	6.7e-11	EGF-like domain
251	pentaxin	1/1	605-743	80.5	2.7e-22	Pentaxin family
251	sushi	1/31	766-820	47.3	9.8e-11	Sushi domain (SCR repeat)
251	sushi	2/31	825-878	68.8	1.2e-16	Sushi domain (SCR repeat)
251	EGF	7/10	884-918	30.0	8.8e-07	EGF-like domain
251	sushi	3/31	924-977	62.9	7e-15	Sushi domain (SCR repeat)
251	sushi	4/31	982-1035	58.5	1.1e-13	Sushi domain (SCR repeat)
251	sushi	5/31	1040-1093	57.5	2.1e-13	Sushi domain (SCR repeat)
251	sushi	6/31	1098-1151	56.3	4.1e-13	Sushi domain (SCR repeat)
251	sushi	7/31	1156-1213	30.6	2.5e-06	Sushi domain (SCR repeat)
251	sushi	8/31	1218-1276	39.4	1.2e-08	Sushi domain (SCR repeat)
251	sushi	9/31	1281-1334	71.9	1.3e-17	Sushi domain (SCR repeat)
251	sushi	10/31	1339-1391	48.3	5.2e-11	Sushi domain (SCR repeat)
251	sushi	11/31	1399-1453	67.3	3.3e-16	Sushi domain (SCR repeat)
251	sushi	12/31	1458-1511	38.9	1.5e-08	Sushi domain (SCR repeat)
251	sushi	13/31	1516-1570	56.2	4.3e-13	Sushi domain (SCR repeat)
251	sushi	14/31	1575-1628	48.6	4.3e-11	Sushi domain (SCR repeat)
251	sushi	15/31	1633-1686	62.1	1.2e-14	Sushi domain (SCR repeat)
251	sushi	16/31	1691-1743	53.8	1.9e-12	Sushi domain (SCR repeat)
251	sushi	18/31	1795-1847	51.8	6.4e-12	Sushi domain (SCR repeat)
251	sushi	19/31	1852-1905	44.0	7.2e-10	Sushi domain (SCR repeat)
251	sushi	20/31	1910-	58.2	1.4e-13	Sushi domain (SCR repeat)

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
			1963			
251	sushi	21/31	1968-2021	60.4	3.4e-14	Sushi domain (SCR repeat)
251	sushi	22/31	2026-2079	51.0	1.1e-11	Sushi domain (SCR repeat)
251	sushi	23/31	2084-2137	54.3	1.4e-12	Sushi domain (SCR repeat)
251	sushi	24/31	2142-2194	38.7	1.8e-08	Sushi domain (SCR repeat)
251	sushi	25/31	2199-2252	48.1	6.2e-11	Sushi domain (SCR repeat)
251	sushi	26/31	2257-2311	47.1	1.1e-10	Sushi domain (SCR repeat)
251	sushi	27/31	2316-2365	31.4	1.5e-06	Sushi domain (SCR repeat)
251	sushi	28/31	2376-2429	53.7	2e-12	Sushi domain (SCR repeat)
251	sushi	29/31	2434-2487	46.6	1.5e-10	Sushi domain (SCR repeat)
251	sushi	30/31	2492-2546	42.1	2.3e-09	Sushi domain (SCR repeat)
251	sushi	31/31	2551-2603	53.3	2.6e-12	Sushi domain (SCR repeat)
251	EGF	8/10	2603-2634	21.6	0.00021	EGF-like domain
251	EGF	9/10	2639-2666	29.8	9.9e-07	EGF-like domain
251	EGF	10/10	2671-2698	22.5	0.00012	EGF-like domain
252	jmjC	1/1	174-288	140.4	1.2e-39	jmjC domain
254	DUF349	1/1	428-444	8.7	0.84	Domain of Unknown Function (DUF349)
255	PSI	1/1	327-372	23.1	3.1e-06	Plexin repeat
255	Glypican	1/1	432-467	4.6	0.98	Glypican
256	DUF279	1/1	68-196	165.9	2.4e-46	Eukaryotic protein of unknown function, D
257	DUF323	1/1	87-342	382.8	3.5e-111	Domain of unknown function (DUF323)
258	lectin_c	1/1	53-164	127.9	1.8e-34	Lectin C-type domain
259	ARD	1/1	3-157	283.0	3.9e-81	ARD/ARD' family
260	Metallophos	1/1	70-285	50.3	3.6e-12	Calcineurin-like phosphoesterase
261	Reprolysin	1/1	218-286	19.3	0.00084	Reprolysin (M12B) family zinc metallo
261	tsp_1	1/8	388-438	48.2	6.8e-13	Thrombospondin type 1 domain
261	tsp_1	7/8	1023-1047	8.5	0.47	Thrombospondin type 1 domain
261	tsp_1	8/8	1079-1102	12.1	0.04	Thrombospondin type 1 domain
263	ig	2/4	171-224	14.6	0.022	Immunoglobulin domain
265	ig	2/4	171-224	14.6	0.022	Immunoglobulin domain
266	ig	2/4	171-224	14.6	0.022	Immunoglobulin domain
267	ig	2/4	185-238	14.6	0.022	Immunoglobulin domain
268	ig	1/1	53-115	23.4	7.9e-05	Immunoglobulin domain
270	AdoHcyase_NAD	1/1	228-389 310.9	1.5e-89 S-adenos		

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
				yl-L-homocysteine hydrolase, NA		
270	AdoHcyase	1/1	41-468	373.5	1.3e-111	S-adenosyl-L-homocysteine hydrolase
271	ig	1/4	34-117	33.9	9.9e-08	Immunoglobulin domain
271	ig	2/4	164-229	22.0	0.00019	Immunoglobulin domain
271	ig	4/4	387-454	36.4	2e-08	Immunoglobulin domain
272	GLTT	1/1	25-53	8.0	0.33	GLTT repeat (6 copies)
273	pentaxin	1/1	342-519	107.1	5.3e-30	Pentaxin family
275	fn3	1/6	39-102	13.8	0.016	Fibronectin type III domain
275	vwa	1/1	186-358	208.8	3.2e-60	von Willebrand factor type A domain
275	fn3	2/6	384-467	52.5	1.1e-13	Fibronectin type III domain
275	fn3	3/6	474-552	65.1	2.5e-17	Fibronectin type III domain
275	fn3	4/6	564-646	31.0	1.7e-07	Fibronectin type III domain
275	fn3	5/6	654-734	46.6	5.4e-12	Fibronectin type III domain
275	fn3	6/6	747-827	59.1	1.3e-15	Fibronectin type III domain
275	TSPN	1/1	849-1044	129.2	1.4e-36	Thrombospondin N-terminal -like domain
275	Collagen	1/3	1079-1122	34.1	6.8e-08	Collagen triple helix repeat (20 copie
275	Collagen	2/3	1124-1180	52.4	6.7e-13	Collagen triple helix repeat (20 copie
276	fn3	1/6	39-102	13.8	0.016	Fibronectin type III domain
276	vwa	1/1	186-358	208.8	3.2e-60	von Willebrand factor type A domain
276	fn3	2/6	384-467	52.5	1.1e-13	Fibronectin type III domain
276	fn3	3/6	474-552	65.1	2.5e-17	Fibronectin type III domain
276	fn3	4/6	564-646	31.0	1.7e-07	Fibronectin type III domain
276	fn3	5/6	654-734	46.6	5.4e-12	Fibronectin type III domain
276	fn3	6/6	747-827	59.1	1.3e-15	Fibronectin type III domain
276	TSPN	1/1	849-1044	129.2	1.4e-36	Thrombospondin N-terminal -like domain
276	Collagen	1/4	1078-1132	31.8	2.9e-07	Collagen triple helix repeat (20 copie
276	Collagen	2/4	1134-1173	26.9	6.5e-06	Collagen triple helix repeat (20 copie
276	Collagen	3/4	1174-1230	52.4	6.7e-13	Collagen triple helix repeat (20 copie
277	fn3	1/6	39-102	13.8	0.016	Fibronectin type III domain
277	vwa	1/1	186-358	208.8	3.2e-60	von Willebrand factor type A domain
277	fn3	2/6	384-467	52.5	1.1e-13	Fibronectin type III domain
277	fn3	3/6	474-552	65.1	2.5e-17	Fibronectin type III domain
277	fn3	4/6	564-646	31.0	1.7e-07	Fibronectin type III domain
277	fn3	5/6	654-734	46.6	5.4e-12	Fibronectin type III domain
277	fn3	6/6	747-827	59.1	1.3e-15	Fibronectin type III domain
277	TSPN	1/1	849-1044	129.2	1.4e-36	Thrombospondin N-terminal -like domain
277	Collagen	1/3	1078-1135	43.2	2.2e-10	Collagen triple helix repeat (20 copie
277	Collagen	2/3	1142-1198	52.4	6.7e-13	Collagen triple helix repeat (20 copie
279	LRR	2/4	89-112	9.8	0.66	Leucine Rich Repeat
279	LRR	3/4	113-136	12.2	0.14	Leucine Rich Repeat

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
279	LRR	4/4	137-160	21.3	0.00036	Leucine Rich Repeat
279	LRRCT	1/1	170-219	44.2	1.5e-13	Leucine rich repeat C-terminal domain
279	EPTP	1/2	223-358	137.3	2.7e-37	EPTP domain
279	EPTP	2/2	411-540	145.1	1.2e-39	EPTP domain
281	7tm_1	1/2	86-124	8.2	0.045	7 transmembrane receptor (rhodopsin family)
282	LRRNT	1/3	73-102	29.6	2e-06	Leucine rich repeat N-terminal domain
282	LRR	1/22	104-127	12.6	0.11	Leucine Rich Repeat
282	LRR	2/22	128-151	15.6	0.015	Leucine Rich Repeat
282	LRR	3/22	152-175	14.5	0.03	Leucine Rich Repeat
282	LRR	4/22	176-199	17.6	0.0041	Leucine Rich Repeat
282	LRR	5/22	200-223	11.9	0.16	Leucine Rich Repeat
282	LRR	6/22	224-247	19.7	0.001	Leucine Rich Repeat
282	LRR	7/22	248-271	12.5	0.11	Leucine Rich Repeat
282	LRR	9/22	296-319	10.0	0.59	Leucine Rich Repeat
282	LRR	10/22	320-341	15.1	0.02	Leucine Rich Repeat
282	LRRCT	1/2	351-399	19.3	3.4e-05	Leucine rich repeat C-terminal domain
282	LRRNT	2/3	436-465	18.3	0.0028	Leucine rich repeat N-terminal domain
282	LRR	13/22	491-514	14.5	0.03	Leucine Rich Repeat
282	LRR	14/22	515-538	13.3	0.067	Leucine Rich Repeat
282	LRR	17/22	587-610	10.6	0.38	Leucine Rich Repeat
282	LRR	18/22	611-634	17.9	0.0032	Leucine Rich Repeat
282	LRR	19/22	635-658	9.4	0.87	Leucine Rich Repeat
282	LRR	20/22	660-683	13.9	0.046	Leucine Rich Repeat
282	LRR	21/22	685-706	12.3	0.13	Leucine Rich Repeat
282	LRRCT	2/2	716-764	51.5	5.5e-16	Leucine rich repeat C-terminal domain
283	SCAN	1/1	45-140	192.7	5.7e-54	SCAN domain
283	zf-C2H2	2/8	283-305	30.2	1.9e-05	Zinc finger, C2H2 type
283	zf-C2H2	3/8	311-333	33.2	3.4e-06	Zinc finger, C2H2 type
283	zf-C2H2	4/8	339-361	25.2	0.00034	Zinc finger, C2H2 type
283	zf-C2H2	5/8	367-389	26.7	0.00014	Zinc finger, C2H2 type
283	zf-C2H2	6/8	395-417	29.6	2.7e-05	Zinc finger, C2H2 type
283	zf-C2H2	7/8	423-445	29.8	2.4e-05	Zinc finger, C2H2 type
283	zf-C2H2	8/8	451-473	32.0	6.6e-06	Zinc finger, C2H2 type
284	Pep_M12B_propep	1/1	82-208	77.1	9.2e-21	Reprolysin family propeptide
284	Reprolysin	2/2	325-422	57.0	9.3e-14	Reprolysin (M12B) family zinc metallo
284	tsp_1	1/2	569-591	15.2	0.0046	Thrombospondin type 1 domain
286	Dor1	1/1	32-388	674.6	4.8e-199	Dor1-like family
287	WH2	1/1	727-744	21.2	0.00024	WH2 motif
291	SAM	1/1	135-198	42.4	1.6e-10	SAM domain (Sterile alpha motif)
292	E1-E2_ATPase	1/1	126-164	8.6	0.13	E1-E2 ATPase
292	Hydrolase	1/2	401-747	27.6	3.4e-06	haloacid dehalogenase-like hydrolase
292	Hydrolase	2/2	816-842	10.9	0.11	haloacid dehalogenase-like hydrolase
293	C2	1/2	12-64	25.7	5.8e-06	C2 domain
293	C2	2/2	112-195	53.9	4.1e-14	C2 domain
294	vwd	1/4	365-521	191.2	1.1e-53	von Willebrand factor type D domain
294	TIL	1/3	640-693	56.8	1.7e-16	Trypsin Inhibitor like cysteine rich d
294	vwd	2/4	756-909	137.5	3e-38	von Willebrand factor type D domain
294	TIL	2/3	1027-1079	47.8	1.1e-13	Trypsin Inhibitor like cysteine rich d

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
294	vwd	3/4	1143-1301	157.1	7e-44	von Willebrand factor type D domain
294	TIL	3/3	1415-1468	45.7	5e-13	Trypsin Inhibitor like cysteine rich d
294	vwd	4/4	1530-1682	151.0	4e-42	von Willebrand factor type D domain
294	zona_pellucida	1/1	1848-2102	330.6	1.7e-95	Zona pellucida-like domain
295	EGF	1/16	147-183	29.1	1.6e-06	EGF-like domain
295	EGF	2/16	190-221	38.0	5.1e-09	EGF-like domain
295	EGF	3/16	228-259	30.3	7.3e-07	EGF-like domain
295	EGF	4/16	266-297	43.5	1.4e-10	EGF-like domain
295	EGF	5/16	308-339	42.6	2.5e-10	EGF-like domain
295	EGF	6/16	344-374	10.9	0.23	EGF-like domain
295	EGF	7/16	383-407	11.6	0.14	EGF-like domain
295	EGF	8/16	420-451	35.3	2.8e-08	EGF-like domain
295	EGF	9/16	459-490	30.2	7.6e-07	EGF-like domain
295	EGF	10/16	498-529	41.8	4.2e-10	EGF-like domain
295	EGF	11/16	536-567	31.9	2.6e-07	EGF-like domain
295	sushi	1/1	573-626	28.7	7.6e-06	Sushi domain (SCR repeat)
295	EGF	12/16	632-663	34.5	4.8e-08	EGF-like domain
295	EGF	13/16	670-701	35.7	2.2e-08	EGF-like domain
295	EGF	14/16	708-739	30.3	7.2e-07	EGF-like domain
295	EGF	15/16	746-777	29.0	1.7e-06	EGF-like domain
295	fn3	1/3	781-862	38.6	1.1e-09	Fibronectin type III domain
295	fn3	2/3	880-963	42.8	6.6e-11	Fibronectin type III domain
295	fn3	3/3	979-1061	45.7	1e-11	Fibronectin type III domain
295	EGF	16/16	1186-1217	38.7	3.1e-09	EGF-like domain
297	zf-C3HC4	1/1	325-365	34.6	2.3e-09	Zinc finger, C3HC4 type (RING finger)
302	zf-C2H2	1/3	86-110	27.6	8.1e-05	Zinc finger, C2H2 type
302	zf-C2H2	2/3	116-140	32.6	4.7e-06	Zinc finger, C2H2 type
302	zf-C2H2	3/3	146-168	29.5	2.8e-05	Zinc finger, C2H2 type
304	ig	1/1	183-237	28.6	3e-06	Immunoglobulin domain
306	pkinase	1/2	39-212	202.4	7.1e-57	Protein kinase domain
306	DUF244	1/1	284-313	4.9	0.69	Uncharacterized protein family (ORF7) DUF
306	pkinase	2/2	276-324	13.4	0.013	Protein kinase domain
307	7tm_1	1/1	58-303	265.6	2.1e-85	7 transmembrane receptor (rhodopsin family)
308	lectin_c	1/1	135-158	10.8	0.24	Lectin C-type domain
309	lectin_c	1/1	135-158	10.8	0.24	Lectin C-type domain
311	lectin_c	1/1	135-158	10.8	0.24	Lectin C-type domain
312	ank	1/5	48-80	39.0	2.6e-09	Ankyrin repeat
312	ank	2/5	111-143	36.6	1.3e-08	Ankyrin repeat
312	ank	3/5	144-166	15.4	0.013	Ankyrin repeat
312	ank	4/5	185-217	46.5	1.9e-11	Ankyrin repeat
312	ank	5/5	220-249	26.4	1e-05	Ankyrin repeat
312	SH3	1/1	298-337	14.6	0.023	SH3 domain
312	SAM	1/2	492-555	74.6	1.1e-19	SAM domain (Sterile alpha motif)
312	SAM	2/2	726-780	57.8	6.5e-15	SAM domain (Sterile alpha motif)
313	LRRNT	1/1	23-49	14.9	0.025	Leucine rich repeat N-terminal domain
313	LRR	1/5	51-74	18.6	0.002	Leucine Rich Repeat
313	LRR	2/5	75-98	18.5	0.0022	Leucine Rich Repeat
313	LRR	3/5	99-122	13.5	0.057	Leucine Rich Repeat

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
313	LRR	4/5	123-146	21.9	0.00024	Leucine Rich Repeat
313	LRRCT	1/1	156-208	47.5	1.2e-14	Leucine rich repeat C-terminal domain
313	ig	1/4	224-283	33.5	1.3e-07	Immunoglobulin domain
313	ig	2/4	320-376	37.7	8.5e-09	Immunoglobulin domain
313	ig	3/4	416-466	22.3	0.00016	Immunoglobulin domain
313	ig	4/4	501-558	32.7	2.1e-07	Immunoglobulin domain
313	An_peroxidase	1/1	702-1241	657.1	9.1e-194	Animal haem peroxidase
313	TILa	1/1	1370-1409	16.9	0.0017	TILa domain
313	vwc	1/1	1371-1426	38.0	1.2e-09	von Willebrand factor type C domain
314	TPR	1/2	82-115	27.7	4.4e-06	TPR Domain
314	TPR	2/2	116-138	11.8	0.15	TPR Domain
314	zf-CCCH	1/4	494-503	8.3	0.94	Zinc finger C-x8-C-x5-C-x3-H type (and simil
314	zf-CCCH	2/4	625-637	8.9	0.61	Zinc finger C-x8-C-x5-C-x3-H type (and simil
314	zf-CCCH	3/4	755-781	18.0	0.0011	Zinc finger C-x8-C-x5-C-x3-H type (and simil
314	zf-C2H2	1/1	842-866	14.9	0.12	Zinc finger, C2H2 type
314	zf-CCCH	4/4	887-913	22.8	3.7e-05	Zinc finger C-x8-C-x5-C-x3-H type (and simil
316	ig	1/3	71-150	22.9	0.00011	Immunoglobulin domain
316	ig	3/3	284-340	15.2	0.015	Immunoglobulin domain
319	FG-GAP	1/5	46-88	21.4	0.00024	FG-GAP repeat
319	FG-GAP	2/5	105-147	21.9	0.00017	FG-GAP repeat
319	FG-GAP	3/5	283-333	23.7	5.4e-05	FG-GAP repeat
319	FG-GAP	5/5	395-437	20.9	0.00033	FG-GAP repeat
320	IRF	1/1	1-76	200.9	1.9e-56	Interferon regulatory factor transcription f
321	ART	1/1	56-291	180.8	2.2e-50	NAD:arginine ADP-ribosyltransferase
322	C1q	1/1	998-1123	101.5	1.7e-26	C1q domain
323	ank	1/1	16-48	33.0	1.3e-07	Ankyrin repeat
324	PRA1	1/1	8-55	15.3	0.0047	Prenylated rab acceptor (PRA1)
327	thiored	1/1	3-64	34.0	6.6e-09	Thioredoxin
328	mito_carr	1/3	9-106	117.2	3e-31	Mitochondrial carrier protein
328	mito_carr	2/3	109-203	114.8	1.6e-30	Mitochondrial carrier protein
328	mito_carr	3/3	208-300	95.4	1.1e-24	Mitochondrial carrier protein
329	EF1BD	1/1	176-262	187.5	2.9e-53	EF-1 guanine nucleotide exchange domain
331	lipocalin	1/1	38-183	119.7	1.4e-33	Lipocalin / cytosolic fatty-acid binding pr
332	MCR_beta	1/1	29-43	4.7	0.97	Methyl-coenzyme M reductase beta subunit, C-
333	cytochrome_c	1/1	2-103	138.2	2.2e-41	Cytochrome c
336	ig	1/5	38-115	29.8	1.3e-06	Immunoglobulin domain
336	ig	2/5	154-210	46.3	3.6e-11	Immunoglobulin domain
336	ig	3/5	243-305	31.9	3.6e-07	Immunoglobulin domain
336	ig	4/5	339-399	19.6	0.00092	Immunoglobulin domain
336	ig	5/5	435-490	25.9	1.6e-05	Immunoglobulin domain
336	fn3	1/2	510-598	19.8	0.0003	Fibronectin type III domain
336	fn3	2/2	619-702	20.0	0.00025	Fibronectin type III domain
337	DUF81	1/1	288-326	10.2	0.099	Domain of unknown function DUF81
338	spectrin	1/7	59-121	15.0	0.011	Spectrin repeat

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
338	spectrin	2/7	124-226	22.2	0.00011	Spectrin repeat
338	spectrin	3/7	229-340	25.7	1.2e-05	Spectrin repeat
338	spectrin	4/7	343-449	19.8	0.00052	Spectrin repeat
338	spectrin	5/7	452-538	23.7	4.2e-05	Spectrin repeat
338	SAA_proteins	1/1	843-860	6.0	0.67	Serum amyloid A protein
338	spectrin	6/7	758-865	47.2	1.3e-11	Spectrin repeat
340	UPF0073	1/1	130-367	427.3	1.4e-124	Uncharacterised protein family (Hly-II
341	Pep_M12B_propep	1/1	33-148	174.6	1.1e-48	Reprolysin family propeptide
341	Reprolysin	1/1	158-355	342.2	5.9e-99	Reprolysin (M12B) family zinc metallo
341	disintegrin	1/1	373-445	30.1	6e-09	Disintegrin
341	DUF38	1/1	471-502	8.2	0.56	Domain of unknown function DUF38
341	EGF	2/2	591-617	11.9	0.12	EGF-like domain
342	CaMBD	1/1	448-464	7.8	0.7	Calmodulin binding domain
342	IQ	2/3	470-490	22.4	0.0002	IQ calmodulin-binding motif
342	IQ	3/3	529-549	21.5	0.00038	IQ calmodulin-binding motif
343	Collagen	1/4	2-30	18.9	0.001	Collagen triple helix repeat (20 copies)
343	Collagen	2/4	68-123	28.2	2.9e-06	Collagen triple helix repeat (20 copies)
343	Collagen	3/4	126-146	15.4	0.0095	Collagen triple helix repeat (20 copies)
343	Collagen	4/4	148-177	19.1	0.00092	Collagen triple helix repeat (20 copies)
344	ig	1/2	221-351	11.8	0.13	Immunoglobulin domain
344	pkinase	1/1	549-882	263.5	2.9e-75	Protein kinase domain
345	SCF	1/1	1-283	698.6	7.7e-211	Stem cell factor
347	PAAD_DAPIN	1/1	18-103	41.6	1.2e-10	PAAD/DAPIN/Pyrin domain
347	RNA_helicase	1/1	195-215	7.9	0.36	RNA helicase
348	fibrinogen_C	1/1	240-457	311.1	1.3e-89	Fibrinogen beta and gamma chains, C-term
349	fibrinogen_C	1/1	240-457	315.6	5.7e-91	Fibrinogen beta and gamma chains, C-term
350	LBP_BPI_CETP_C	1/1	290-428	45.8	1.3e-11	LBP / BPI / CETP family, C-terminal do
351	Oxysterol_BP	1/2	19-270	299.0	5.9e-86 Oxysterol-binding protein	
351	Oxysterol_BP	2/2	329-429 45.7	1.1e-11 Oxysterol-binding protein		
352	APC10	2/2	125-152	10.8	0.029	Anaphase-promoting complex, subunit 10
352	Pox_TAA1	1/1	704-717	7.3	0.71	Poxvirus trans-activator protein A1
352	BK_channel_a	1/1	1069-1082	4.3	0.73	Calcium-activated BK potassium channel
352	ZZ	1/2	1598-	26.4	2.4e-05	Zinc finger, ZZ type

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
			1641			
352	ZZ	2/2	1642-1686	32.1	7.1e-07	Zinc finger, ZZ type
353	Collagen	1/2	37-64	18.8	0.0011	Collagen triple helix repeat (20 copies)
353	Collagen	2/2	65-124	48.8	6.4e-12	Collagen triple helix repeat (20 copies)
353	C1q	1/1	134-258	148.4	1.3e-40	C1q domain
355	ion_trans	1/2	70-192	29.5	8.2e-07	Ion transport protein
356	ion_trans	1/2	75-197	29.5	8.2e-07	Ion transport protein
357	gntR	1/1	109-124	7.6	0.91	Bacterial regulatory proteins, gntR family
357	A2M_N	1/1	1-613	310.7	5.4e-91	Alpha-2-macroglobulin family N-terminal regi
357	A2M	1/1	721-1448	711.6	9.2e-214	Alpha-2-macroglobulin family
358	PAX	1/1	4-142	279.7	3.8e-80	'Paired box' domain
358	homeobox	1/1	225-281	112.7	7.1e-30	Homeobox domain
359	Collagen	1/1	41-88	37.2	9.7e-09	Collagen triple helix repeat (20 copies)
359	lectin_c	1/1	135-238	78.4	1.5e-19	Lectin C-type domain
360	Collagen	1/3	24-82	48.3	8.8e-12	Collagen triple helix repeat (20 copies)
360	Collagen	2/3	95-154	42.8	2.9e-10	Collagen triple helix repeat (20 copies)
360	Collagen	3/3	155-191	33.6	9.8e-08	Collagen triple helix repeat (20 copies)
360	C1q	1/1	203-329	150.7	2.6e-41	C1q domain
363	Xlink	1/1	26-52	10.9	0.00037	Extracellular link domain
363	lectin_c	1/1	34-160	70.4	3.7e-17	Lectin C-type domain
369	Collagen	1/1	61-109	34.2	6.4e-08	Collagen triple helix repeat (20 copies)
369	C1q	1/1	128-252	117.4	2.7e-31	C1q domain
371	ig	1/1	42-98	17.8	0.0028	Immunoglobulin domain
374	SH2	1/2	10-87	103.3	1.5e-34	SH2 domain
374	SH2	2/2	163-239	107.5	5.4e-36	SH2 domain
374	pkinase	1/1	338-586	266.4	3.9e-76	Protein kinase domain
375	SCP	1/1	66-205	165.1	1.1e-45	SCP-like extracellular protein
375	LCCL	1/2	293-384	181.6	1.9e-52	LCCL domain
375	LCCL	2/2	394-488	183.7	4.5e-53	LCCL domain
379	CD20	1/1	24-56	15.8	0.0016	CD20/IgE Fc receptor beta subunit family
381	Radical_SAM	1/1	131-296	96.3	5.8e-26	Radical SAM superfamily
383	Peptidase_M10	1/2	23-69	100.6	2.1e-26	Matrixin
383	PG_binding_1	1/1	85-115	10.3	0.28	Putative peptidoglycan binding domain
383	Peptidase_M10_N	1/1	79-120	88.6	4.3e-30	Matrix metalloprotease, N-terminal do
383	Peptidase_M10	2/2	127-231	189.0	7.7e-53	Matrixin
383	Fragilysin	1/1	238-263	9.8	0.054	Fragilysin metallopeptidase (M10C) en
383	hemopexin	2/3	309-350	46.8	1.3e-12	Hemopexin
384	Collagen	1/10	2-58	42.7	3.1e-10	Collagen triple helix repeat (20 copies)
384	Collagen	2/10	59-118	50.8	1.8e-12	Collagen triple helix repeat (20 copies)
384	Collagen	3/10	122-181	51.9	9.1e-13	Collagen triple helix repeat (20 copies)
384	Collagen	4/10	182-241	40.6	1.1e-09	Collagen triple helix repeat (20 copies)
384	Collagen	5/10	242-301	51.8	9.3e-13	Collagen triple helix repeat (20 copies)
384	Collagen	6/10	303-350	40.4	1.3e-09	Collagen triple helix repeat (20 copies)
384	Collagen	7/10	351-406	40.5	1.2e-09	Collagen triple helix repeat (20 copies)
384	Collagen	8/10	408-462	40.5	1.2e-09	Collagen triple helix repeat (20 copies)

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
384	Collagen	9/10	465-524	38.9	3.3e-09	Collagen triple helix repeat (20 copies)
384	Collagen	10/10	525-584	42.8	2.8e-10	Collagen triple helix repeat (20 copies)
384	COLFI	1/2	639-697	92.7	1.1e-35	Fibrillar collagen C-terminal domain
384	COLFI	2/2	706-822	56.7	1.1e-21	Fibrillar collagen C-terminal domain
387	DUF28	1/1	61-297	156.4	4.8e-43	Domain of unknown function DUF28
392	Spore_permease	1/1	251-281	9.0	0.15	Spore germination protein
392	7tm_1	1/1	68-322	159.7	4.1e-51	7 transmembrane receptor (rhodopsin fa
393	Spore_permease	1/1	234-264	9.0	0.15	Spore germination protein
393	7tm_1	1/1	51-305	159.7	4.1e-51	7 transmembrane receptor (rhodopsin fa
395	FCH	1/1	14-102	81.3	5.3e-22	Fes/CIP4 homology domain
395	SH3	1/1	366-422	70.1	1.2e-17	SH3 domain
396	HSP70	1/1	3-380	364.0	1.1e-105	Hsp70 protein
397	ig	2/5	150-207	24.1	5.1e-05	Immunoglobulin domain
397	ig	3/5	242-291	24.1	5.2e-05	Immunoglobulin domain
397	ig	4/5	367-385	13.2	0.055	Immunoglobulin domain
398	ig	2/3	149-206	24.1	5.1e-05	Immunoglobulin domain
398	ig	3/3	241-290	24.1	5.2e-05	Immunoglobulin domain
398	PPTA	1/1	324-336	8.6	1	Protein prenyltransferase alpha subunit repe
399	ig	2/3	255-312	24.1	5.1e-05	Immunoglobulin domain
399	ig	3/3	347-396	24.1	5.2e-05	Immunoglobulin domain
399	PPTA	1/1	430-442	8.6	1	Protein prenyltransferase alpha subunit repe
400	Pep_M12B_propep	1/1	75-191	106.1	4.7e-29	Reprolysin family propeptide
400	Reprolysin	1/1	341-370	22.8	0.0001	Reprolysin (M12B) family zinc metallo
400	disintegrin	1/1	419-494	48.9	3.4e-15	Disintegrin
401	Pep_M12B_propep	1/1	75-191	104.6	1.2e-28	Reprolysin family propeptide
402	serpin	1/1	47-415	753.0	1.2e-222	Serpin (serine protease inhibitor)
403	KRAB	1/1	39-79	89.1	9.5e-24	KRAB box
403	zf-C2H2	1/16	204-223	27.2	0.0001	Zinc finger, C2H2 type
403	zf-C2H2	2/16	232-254	30.5	1.6e-05	Zinc finger, C2H2 type
403	zf-C2H2	3/16	260-282	24.3	0.00054	Zinc finger, C2H2 type
403	zf-C2H2	4/16	288-310	27.4	9.4e-05	Zinc finger, C2H2 type
403	zf-C2H2	5/16	316-338	17.0	0.036	Zinc finger, C2H2 type
403	zf-C2H2	6/16	344-366	28.2	5.8e-05	Zinc finger, C2H2 type
403	zf-C2H2	7/16	372-394	18.1	0.019	Zinc finger, C2H2 type
403	zf-C2H2	8/16	400-422	25.9	0.00022	Zinc finger, C2H2 type
403	zf-C2H2	9/16	428-450	29.7	2.4e-05	Zinc finger, C2H2 type
403	zf-C2H2	10/16	456-478	33.8	2.4e-06	Zinc finger, C2H2 type
403	zf-C2H2	11/16	484-505	19.2	0.01	Zinc finger, C2H2 type
403	zf-C2H2	12/16	511-533	25.4	0.00028	Zinc finger, C2H2 type
403	zf-C2H2	13/16	539-561	34.3	1.8e-06	Zinc finger, C2H2 type
403	zf-C2H2	14/16	567-589	24.8	0.00041	Zinc finger, C2H2 type
403	zf-C2H2	15/16	595-617	21.5	0.0028	Zinc finger, C2H2 type
403	zf-C2H2	16/16	623-645	34.5	1.6e-06	Zinc finger, C2H2 type
404	CLP_protease	1/2	67-106	57.7	1.3e-14	Clp protease
404	CLP_prot	2/2	107-197	152.3	8.6e-42	Clp protease

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
	ease					
408	zf-C2H2	1/1	174-196	18.9	0.012	Zinc finger, C2H2 type
410	F-box	1/1	131-171	13.0	0.11	F-box domain
411	Collagen	1/3	2-19	10.0	0.29	Collagen triple helix repeat (20 copies)
411	Collagen	2/3	36-84	39.1	2.9e-09	Collagen triple helix repeat (20 copies)
411	Collagen	3/3	87-146	50.3	2.5e-12	Collagen triple helix repeat (20 copies)
412	EGF	1/8	129-165	20.8	0.00037	EGF-like domain
412	EGF	2/8	169-204	21.3	0.00026	EGF-like domain
412	EGF	3/8	238-273	28.9	1.8e-06	EGF-like domain
412	EGF	4/8	279-314	25.4	1.8e-05	EGF-like domain
412	EGF	5/8	320-353	14.3	0.025	EGF-like domain
412	EGF	6/8	372-407	29.5	1.3e-06	EGF-like domain
412	TNFR_c6	1/3	655-672	12.1	0.034	TNFR/NGFR cysteine-rich region
412	TNFR_c6	2/3	759-780	9.6	0.21	TNFR/NGFR cysteine-rich region
412	CUB	1/2	870-908	52.5	3.3e-14	CUB domain
412	CUB	2/2	947-979	18.4	0.00036	CUB domain
413	EGF	1/8	3-39	20.8	0.00037	EGF-like domain
413	EGF	2/8	43-78	21.3	0.00026	EGF-like domain
413	EGF	3/8	112-147	28.9	1.8e-06	EGF-like domain
413	EGF	4/8	153-188	25.4	1.8e-05	EGF-like domain
413	EGF	5/8	194-227	14.3	0.025	EGF-like domain
413	EGF	6/8	246-281	29.5	1.3e-06	EGF-like domain
413	TNFR_c6	1/3	529-546	12.1	0.034	TNFR/NGFR cysteine-rich region
413	TNFR_c6	2/3	633-654	9.6	0.21	TNFR/NGFR cysteine-rich region
413	CUB	1/2	744-782	52.5	3.3e-14	CUB domain
413	CUB	2/2	821-853	18.4	0.00036	CUB domain
414	COX6C	1/1	1-75	139.9	2.5e-42	Cytochrome c oxidase subunit VIc
415	ig	1/2	39-97	15.6	0.012	Immunoglobulin domain
415	ig	2/2	128-189	44.6	1.1e-10	Immunoglobulin domain
417	ig	3/3	153-206	20.1	0.00067	Immunoglobulin domain
418	PP2C	1/1	128-172	8.1	0.26	Protein phosphatase 2C
419	ig	3/3	253-302	31.6	4.4e-07	Immunoglobulin domain
421	UPAR_LY6	2/2	124-138	12.5	0.44	u-PAR/Ly-6 domain
423	SCP	1/1	52-181	124.5	5.2e-34	SCP-like extracellular protein
423	EGF	1/2	225-260	15.7	0.0098	EGF-like domain
424	ig	1/1	55-144	26.7	9.8e-06	Immunoglobulin domain
425	7tm_1	1/1	2-219	85.7	3.6e-27	7 transmembrane receptor (rhodopsin family)
426	Activin_recp	1/1	98-112	5.9	0.76	Activin types I and II receptor domain
432	toxin	1/1	82-96	10.9	0.47	Snake toxin
432	UPAR_LY6	1/1	23-96	33.6	4.6e-06	u-PAR/Ly-6 domain
432	Activin_recp	1/1	83-97	6.2	0.61	Activin types I and II receptor domain
435	Peptidase_C54	1/2	109-168	119.9	2.4e-38	Peptidase family C54
435	Peptidase_C54	2/2	210-407	267.4	2e-86	Peptidase family C54
436	ig	1/4	85-121	10.2	0.37	Immunoglobulin domain
436	ig	2/4	162-219	11.9	0.12	Immunoglobulin domain
436	ig	3/4	255-312	16.5	0.0066	Immunoglobulin domain
436	ig	4/4	347-396	32.3	2.8e-07	Immunoglobulin domain
437	ig	1/3	85-121	9.0	0.8	Immunoglobulin domain

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
437	ig	2/3	162-219	19.2	0.0012	Immunoglobulin domain
437	ig	3/3	254-303	30.2	1.1e-06	Immunoglobulin domain
438	ig	1/3	107-143	10.1	0.39	Immunoglobulin domain
438	ig	3/3	277-334	16.0	0.0089	Immunoglobulin domain
439	tsp_1	1/3	37-81	25.9	3e-06	Thrombospondin type 1 domain
439	tsp_1	3/3	363-387	17.4	0.0011	Thrombospondin type 1 domain
440	tsp_1	1/7	37-81	25.9	3e-06	Thrombospondin type 1 domain
440	tsp_1	3/7	380-404	17.4	0.0011	Thrombospondin type 1 domain
440	tsp_1	4/7	444-463	21.1	8.3e-05	Thrombospondin type 1 domain
440	tsp_1	5/7	531-550	19.8	0.0002	Thrombospondin type 1 domain
441	tsp_1	1/7	85-129	25.9	3e-06	Thrombospondin type 1 domain
441	tsp_1	3/7	428-452	17.4	0.0011	Thrombospondin type 1 domain
441	tsp_1	4/7	492-511	21.1	8.3e-05	Thrombospondin type 1 domain
441	tsp_1	5/7	579-598	19.8	0.0002	Thrombospondin type 1 domain
442	UPAR_LY6	1/1	23-101	33.2	5.9e-06	u-PAR/Ly-6 domain
443	UPAR_LY6	1/1	21-94	87.2	3.3e-22	u-PAR/Ly-6 domain
443	Activin_recp	1/1	86-100	7.5	0.25	Activin types I and II receptor domain
444	UPAR_LY6	1/1	21-55	34.8	2e-06	u-PAR/Ly-6 domain
446	LRRNT	1/1	33-60	31.2	7e-07	Leucine rich repeat N-terminal domain
446	LRR	2/10	86-109	17.8	0.0036	Leucine Rich Repeat
446	LRR	3/10	110-133	11.2	0.26	Leucine Rich Repeat
446	LRR	4/10	134-157	19.5	0.0012	Leucine Rich Repeat
446	LRR	5/10	158-181	14.6	0.028	Leucine Rich Repeat
446	LRR	6/10	182-205	17.8	0.0035	Leucine Rich Repeat
446	LRR	7/10	206-229	12.4	0.12	Leucine Rich Repeat
446	LRR	9/10	254-275	13.0	0.083	Leucine Rich Repeat
446	LRR	10/10	279-302	12.1	0.15	Leucine Rich Repeat
446	LRRCT	1/1	312-362	16.3	0.00033	Leucine rich repeat C-terminal domain
447	ig	1/2	159-217	24.5	4.1e-05	Immunoglobulin domain
447	ig	2/2	267-321	25.3	2.4e-05	Immunoglobulin domain
448	Collagen	1/17	1-55	45.4	5.3e-11	Collagen triple helix repeat (20 copies)
448	Collagen	2/17	56-115	75.7	2.5e-19	Collagen triple helix repeat (20 copies)
448	Collagen	3/17	116-175	64.9	2.4e-16	Collagen triple helix repeat (20 copies)
448	Collagen	4/17	176-235	61.6	1.9e-15	Collagen triple helix repeat (20 copies)
448	Collagen	5/17	236-295	61.1	2.6e-15	Collagen triple helix repeat (20 copies)
448	Collagen	6/17	296-355	63.9	4.4e-16	Collagen triple helix repeat (20 copies)
448	Collagen	7/17	356-415	64.6	2.9e-16	Collagen triple helix repeat (20 copies)
448	Collagen	8/17	416-475	62.1	1.4e-15	Collagen triple helix repeat (20 copies)
448	Collagen	9/17	476-535	60.6	3.6e-15	Collagen triple helix repeat (20 copies)
448	Collagen	10/17	536-595	70.2	8.4e-18	Collagen triple helix repeat (20 copies)
448	Collagen	11/17	599-658	68.4	2.7e-17	Collagen triple helix repeat (20 copies)
448	Collagen	12/17	659-718	60.4	4e-15	Collagen triple helix repeat (20 copies)
448	Collagen	13/17	719-778	59.2	8.9e-15	Collagen triple helix repeat (20 copies)
448	Collagen	14/17	779-838	62.6	9.9e-16	Collagen triple helix repeat (20 copies)
448	Collagen	15/17	839-898	60.1	5.1e-15	Collagen triple helix repeat (20 copies)
448	Collagen	16/17	899-958	74.1	7.2e-19	Collagen triple helix repeat (20 copies)
448	Collagen	17/17	959-1012	40.5	1.2e-09	Collagen triple helix repeat (20 copies)
448	COLFI	1/1	1065-1283	565.2	2.2e-220	Fibrillar collagen C-terminal domain
449	IL1	2/2	62-157	75.6	4e-20	Interleukin-1 / 18
450	trypsin	1/1	56-101	69.8	2.5e-21	Trypsin

TABLE 3A

SEQ ID	Model	Repeats	Position	Score	E value	Description
451	trypsin	1/1	28-262	250.0	1.1e-78	Trypsin
453	Collagen	1/11	77-101	14.9	0.013	Collagen triple helix repeat (20 copies)
453	Collagen	3/11	126-168	34.9	4.3e-08	Collagen triple helix repeat (20 copies)
453	Collagen	4/11	173-209	29.3	1.4e-06	Collagen triple helix repeat (20 copies)
453	Collagen	5/11	211-235	8.3	0.83	Collagen triple helix repeat (20 copies)
453	Collagen	6/11	237-280	32.2	2.3e-07	Collagen triple helix repeat (20 copies)
453	Collagen	7/11	281-314	22.7	9.6e-05	Collagen triple helix repeat (20 copies)
453	Collagen	8/11	316-375	45.9	3.9e-11	Collagen triple helix repeat (20 copies)
453	Collagen	9/11	376-430	41.4	6.7e-10	Collagen triple helix repeat (20 copies)
453	Collagen	10/11	433-492	44.9	7.6e-11	Collagen triple helix repeat (20 copies)
453	Collagen	11/11	495-535	30.3	7.8e-07	Collagen triple helix repeat (20 copies)
453	C1q	1/1	576-700	263.2	3.4e-75	C1q domain
455	Transposase_22	1/1	2-28	11.7	0.0042	L1 transposable element
456	Ribosomal_S28e	1/1	57-97	41.9	1.2e-11	Ribosomal protein S28e
457	LRR	2/10	73-96	11.0	0.29	Leucine Rich Repeat
457	LRR	3/10	97-120	17.9	0.0033	Leucine Rich Repeat
457	LRR	9/10	444-467	16.4	0.009	Leucine Rich Repeat
457	LRRCT	1/1	522-575	43.9	2e-13	Leucine rich repeat C-terminal domain
457	TIR	1/1	636-774	113.5	4e-33	TIR domain
460	UPAR_LY6	1/1	23-101	30.8	3.2e-05	u-PAR/Ly-6 domain
460	Activin_recp	1/1	72-107	7.4	0.27	Activin types I and II receptor domain
461	UPAR_LY6	1/1	123-161	11.7	0.69	u-PAR/Ly-6 domain
462	Pep_M12B_propep	1/1	33-148	174.6	1.1e-48	Reprolysin family propeptide
462	Reprolysin	1/1	158-355	342.2	5.9e-99	Reprolysin (M12B) family zinc metallo
462	disintegrin	2/2	422-477	21.7	3.8e-06	disintegrin
462	DUF38	1/1	503-534	8.2	0.56	Domain of unknown function DUF38
462	EGF	2/2	623-649	11.9	0.12	EGF-like domain
463	Pep_M12B_propep	1/1	33-148	174.6	1.1e-48	Reprolysin family propeptide
463	Reprolysin	1/1	158-329	292.8	4.4e-84	Reprolysin (M12B) family zinc metallo
464	Reprolysin	1/1	41-72	21.2	0.00026	Reprolysin (M12B) family zinc metallop
465	Pep_M12B_propep	1/1	1-83	113.2	4.2e-31	Reprolysin family propeptide
465	Reprolysin	1/1	93-107	18.7	0.0012	Reprolysin (M12B) family zinc metallo

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
237	UCH	1/2	38-319	129.5	1.4e-42	Ubiquitin carboxyl-terminal hydrolase
237	UCH	2/2	448-479	27.4	2.5e-09	Ubiquitin carboxyl-terminal hydrolase
237	DUF706	1/1	1119-1129	6.0	0.088	Family of unknown function (DUF706)
238	ig	1/2	31-89	28.4	9.8e-07	Immunoglobulin domain
238	ig	2/2	126-182	22.0	4.8e-05	Immunoglobulin domain
241	Hormone_1	1/1	9-215	305.7	9e-113	Somatotropin hormone family
242	TSP_1	1/3	16-66	59.7	2.6e-17	Thrombospondin type 1 domain
242	TSP_1	2/3	73-123	41.1	9.7e-12	Thrombospondin type 1 domain
242	TIL	1/6	108-125	0.1	13	Trypsin Inhibitor like cysteine rich do
242	TSP_1	3/3	130-180	54.7	8.3e-16	Thrombospondin type 1 domain
242	G2F	1/1	181-368	359.5	4.3e-105	G2F domain
242	EGF	1/7	403-417	5.4	1.1	EGF-like domain
242	TIL	2/6	403-423	1.0	6.4	Trypsin Inhibitor like cysteine rich do
242	EGF	2/7	423-457	30.7	1.1e-07	EGF-like domain
242	EGF	3/7	463-502	11.9	0.018	EGF-like domain
242	EGF	4/7	508-540	21.8	3.3e-05	EGF-like domain
242	TIL	3/6	527-546	12.5	0.0014	Trypsin Inhibitor like cysteine rich do
242	EGF	5/7	546-567	8.2	0.2	EGF-like domain
242	TIL	4/6	578-588	0.7	8.3	Trypsin Inhibitor like cysteine rich do
242	EGF	6/7	588-625	25.7	2.8e-06	EGF-like domain
242	TIL	5/6	609-631	0.8	7.7	Trypsin Inhibitor like cysteine rich do
242	EGF	7/7	631-665	36.8	2.2e-09	EGF-like domain
242	TIL	6/6	650-671	8.4	0.028	Trypsin Inhibitor like cysteine rich do
245	priB_priC	1/1	676-696	10.6	0.011	Primosomal replication protein priB a
245	Drf_FH1	1/2	856-964	48.8	8.8e-13	Formin Homology Region 1
245	Drf_FH1	2/2	965-1115	116.1	8.4e-32	Formin Homology Region 1
245	FH2	1/1	1141-1530	452.9	3.4e-133	Formin Homology 2 Domain
246	zf-C3HC4	1/1	127-138	5.9	0.053	Zinc finger, C3HC4 type (RING finger)
248	VWA	1/1	83-255	131.4	4.7e-41	von Willebrand factor type A domain
248	EGF	1/13	281-314	2.4	7.8	EGF-like domain
248	Laminin_EGF	1/12	307-320	1.3	9.3	Laminin EGF-like (Domains III and V)
248	TNFR_c6	1/5	307-328	1.3	13	TNFR/NGFR cysteine-rich region
248	EB	1/5	360-373	4.6	0.67	EB module
248	EGF	2/13	360-373	3.2	4.7	EGF-like domain
248	TIL	1/3	360-373	2.5	2.1	Trypsin Inhibitor like cysteine rich do
248	Laminin_EGF	2/12	362-373	1.4	8.3	Laminin EGF-like (Domains III and V)
248	Sushi	1/34	378-433	33.9	8.4e-08	Sushi domain (SCR repeat)
248	Paramecium_SA	1/6	425-439	3.3	0.84	Paramecium surface antigen domain
248	Sushi	2/34	438-493	58.3	2e-14	Sushi domain (SCR repeat)

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
248	Paramecium_SA	2/6	486-499	0.5	6.9	Paramecium surface antigen domain
248	Sushi	3/34	498-559	12.7	0.023	Sushi domain (SCR repeat)
248	HYR	1/2	561-642	68.0	8.6e-19	HYR domain
248	HYR	2/2	644-722	65.3	4.9e-18	HYR domain
248	EGF	3/13	739-749	0.4	28	EGF-like domain
248	EB	2/5	988-999	1.4	7	EB module
248	TNFR_c6	2/5	1002-1017	3.2	3.6	TNFR/NGFR cysteine-rich region
248	TNFR_c6	3/5	1018-1042	11.5	0.014	TNFR/NGFR cysteine-rich region
248	TNFR_c6	4/5	1056-1072	6.5	0.39	TNFR/NGFR cysteine-rich region
248	Laminin_EGF	3/12	1069-1086	0.2	18	Laminin EGF-like (Domains III and V)
248	TNFR_c6	5/5	1110-1126	8.5	0.1	TNFR/NGFR cysteine-rich region
248	EGF	4/13	1197-1228	35.5	5.4e-09	EGF-like domain
248	Laminin_EGF	4/12	1202-1229	2.8	3.4	Laminin EGF-like (Domains III and V)
248	EGF	5/13	1235-1266	45.1	1.2e-11	EGF-like domain
248	EB	3/5	1240-1266	1.1	8.7	EB module
248	Laminin_EGF	5/12	1255-1268	4.7	0.92	Laminin EGF-like (Domains III and V)
248	DSL	1/6	1257-1266	1.0	8.3	Delta serrate ligand
248	EGF	6/13	1273-1304	34.9	7.5e-09	EGF-like domain
248	EB	4/5	1278-1287	0.3	16	EB module
248	Laminin_EGF	6/12	1284-1305	0.3	18	Laminin EGF-like (Domains III and V)
248	EGF	7/13	1311-1342	35.1	6.8e-09	EGF-like domain
248	EB	5/5	1316-1342	4.3	0.84	EB module
248	EGF	8/13	1349-1380	40.4	2.3e-10	EGF-like domain
248	Laminin_EGF	7/12	1360-1381	8.0	0.1	Laminin EGF-like (Domains III and V)
248	DSL	2/6	1370-1380	5.9	0.27	Delta serrate ligand
248	EGF	9/13	1387-1418	44.6	1.6e-11	EGF-like domain
248	Laminin_EGF	8/12	1407-1419	7.0	0.2	Laminin EGF-like (Domains III and V)
248	DSL	3/6	1409-1418	1.1	7.6	Delta serrate ligand
248	Pentaxin	1/1	1470-1608	80.5	1.6e-25	Pentaxin family
248	Sushi	4/34	1631-1685	47.3	3.1e-11	Sushi domain (SCR repeat)
248	Sushi	5/34	1690-	68.8	1.4e-17	Sushi domain (SCR repeat)

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
			1743			
248	Paramecium_SA	3/6	1736-1750	5.3	0.19	Paramecium surface antigen domain
248	EGF	10/13	1749-1783	29.9	1.9e-07	EGF-like domain
248	Sushi	6/34	1789-1842	62.9	8.7e-16	Sushi domain (SCR repeat)
248	Sushi	7/34	1847-1900	58.5	1.8e-14	Sushi domain (SCR repeat)
248	Sushi	8/34	1905-1958	57.5	3.7e-14	Sushi domain (SCR repeat)
248	Sushi	9/34	1963-2016	56.3	8.1e-14	Sushi domain (SCR repeat)
248	Sushi	10/34	2021-2078	30.6	6e-07	Sushi domain (SCR repeat)
248	Sushi	11/34	2083-2141	39.4	3.3e-09	Sushi domain (SCR repeat)
248	Sushi	12/34	2146-2199	71.9	1.7e-18	Sushi domain (SCR repeat)
248	Sushi	13/34	2204-2256	48.3	1.7e-11	Sushi domain (SCR repeat)
248	Sushi	14/34	2264-2318	67.3	4.1e-17	Sushi domain (SCR repeat)
248	Sushi	15/34	2323-2376	38.9	4.3e-09	Sushi domain (SCR repeat)
248	Sushi	16/34	2381-2435	56.3	8.5e-14	Sushi domain (SCR repeat)
248	Sushi	17/34	2440-2493	48.6	1.4e-11	Sushi domain (SCR repeat)
248	Paramecium_SA	4/6	2486-2499	0.1	9.7	Paramecium surface antigen domain
248	Sushi	18/34	2498-2551	62.1	1.5e-15	Sushi domain (SCR repeat)
248	Sushi	19/34	2556-2608	53.8	4.7e-13	Sushi domain (SCR repeat)
248	HRM	1/2	2575-2629	8.3	0.12	Hormone receptor domain
248	Sushi	20/34	2613-2625	3.7	4.7	Sushi domain (SCR repeat)
248	Sushi	21/34	2660-2712	51.8	1.9e-12	Sushi domain (SCR repeat)
248	Paramecium_SA	5/6	2704-2718	8.5	0.018	Paramecium surface antigen domain
248	Sushi	22/34	2717-2770	44.0	2.2e-10	Sushi domain (SCR repeat)
248	Sushi	23/34	2775-2828	58.2	2.3e-14	Sushi domain (SCR repeat)
248	Laminin_EGF	9/12	2800-2815	0.5	16	Laminin EGF-like (Domains III and V)
248	TIL	2/3	2800-2815	5.9	0.18	Trypsin Inhibitor like cysteine rich do
248	Sushi	24/34	2833-2886	60.4	4.8e-15	Sushi domain (SCR repeat)
248	Paramecium_SA	6/6	2879-2892	1.2	4.2	Paramecium surface antigen domain
248	Sushi	25/34	2891-	51.0	3.3e-12	Sushi domain (SCR repeat)

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
			2944			
248	Sushi	26/34	2949-3002	54.3	3.3e-13	Sushi domain (SCR repeat)
248	HRM	2/2	2983-2995	5.6	0.69	Hormone receptor domain
248	Sushi	27/34	3007-3059	38.7	5.1e-09	Sushi domain (SCR repeat)
248	TIL	3/3	3046-3066	1.3	5.3	Trypsin Inhibitor like cysteine rich do
248	Sushi	28/34	3064-3117	48.1	2e-11	Sushi domain (SCR repeat)
248	Sushi	29/34	3122-3176	47.1	3.4e-11	Sushi domain (SCR repeat)
248	Laminin_EGF	10/12	3147-3163	2.6	3.7	Laminin EGF-like (Domains III and V)
248	Sushi	30/34	3181-3230	31.4	3.6e-07	Sushi domain (SCR repeat)
248	Sushi	31/34	3241-3294	53.7	5e-13	Sushi domain (SCR repeat)
248	Sushi	32/34	3299-3352	46.6	4.7e-11	Sushi domain (SCR repeat)
248	Sushi	33/34	3357-3411	42.1	6.7e-10	Sushi domain (SCR repeat)
248	Sushi	34/34	3416-3468	53.3	6.6e-13	Sushi domain (SCR repeat)
248	C_tripleX	1/2	3462-3478	6.8	0.17	Cysteine rich repeat
248	EGF	11/13	3468-3499	22.6	1.9e-05	EGF-like domain
248	Laminin_EGF	11/12	3487-3501	1.6	7.2	Laminin EGF-like (Domains III and V)
248	DSL	4/6	3489-3499	5.6	0.32	Delta serrate ligand
248	EGF	12/13	3504-3531	29.9	1.9e-07	EGF-like domain
248	Laminin_EGF	12/12	3509-3531	4.2	1.3	Laminin EGF-like (Domains III and V)
248	DSL	5/6	3522-3531	6.7	0.15	Delta serrate ligand
248	C_tripleX	2/2	3534-3548	1.3	12	Cysteine rich repeat
248	EGF	13/13	3536-3563	22.5	2.1e-05	EGF-like domain
248	DSL	6/6	3554-3563	2.3	3.2	Delta serrate ligand
249	VWA	1/1	83-255	131.4	4.7e-41	von Willebrand factor type A domain
249	Sushi	1/3	378-433	33.9	8.4e-08	Sushi domain (SCR repeat)
249	Sushi	2/3	438-493	58.3	2e-14	Sushi domain (SCR repeat)
249	Sushi	3/3	498-559	12.7	0.023	Sushi domain (SCR repeat)
249	HYR	1/2	561-642	68.0	8.6e-19	HYR domain
249	HYR	2/2	644-722	65.3	4.9e-18	HYR domain
250	TNFR_c6	1/4	137-152	3.2	3.6	TNFR/NGFR cysteine-rich region
250	TNFR_c6	2/4	153-177	11.5	0.014	TNFR/NGFR cysteine-rich region
250	TNFR_c6	3/4	191-207	6.5	0.39	TNFR/NGFR cysteine-rich region
250	TNFR_c6	4/4	245-261	8.5	0.1	TNFR/NGFR cysteine-rich region

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
250	EGF	1/3	332-363	35.5	5.4e-09	EGF-like domain
250	EGF	2/3	370-401	45.1	1.2e-11	EGF-like domain
250	EGF	3/3	408-437	27.3	9.7e-07	EGF-like domain
251	TNFR_c6	1/4	137-152	3.2	3.6	TNFR/NGFR cysteine-rich region
251	TNFR_c6	2/4	153-177	11.5	0.014	TNFR/NGFR cysteine-rich region
251	TNFR_c6	3/4	191-207	6.5	0.39	TNFR/NGFR cysteine-rich region
251	Laminin_EGF	1/10	204-221	0.2	18	Laminin EGF-like (Domains III and V)
251	TNFR_c6	4/4	245-261	8.5	0.1	TNFR/NGFR cysteine-rich region
251	EGF	1/10	332-363	35.5	5.4e-09	EGF-like domain
251	Laminin_EGF	2/10	337-364	2.8	3.4	Laminin EGF-like (Domains III and V)
251	EGF	2/10	370-401	45.1	1.2e-11	EGF-like domain
251	Laminin_EGF	3/10	390-403	4.7	0.92	Laminin EGF-like (Domains III and V)
251	DSL	1/6	392-401	1.0	8.3	Delta serrate ligand
251	EGF	3/10	408-439	34.9	7.5e-09	EGF-like domain
251	Laminin_EGF	4/10	419-440	0.3	18	Laminin EGF-like (Domains III and V)
251	EGF	4/10	446-477	35.1	6.8e-09	EGF-like domain
251	EGF	5/10	484-515	40.4	2.3e-10	EGF-like domain
251	Laminin_EGF	5/10	495-516	8.0	0.1	Laminin EGF-like (Domains III and V)
251	DSL	2/6	505-515	5.9	0.27	Delta serrate ligand
251	EGF	6/10	522-553	44.6	1.6e-11	EGF-like domain
251	Laminin_EGF	6/10	542-554	7.0	0.2	Laminin EGF-like (Domains III and V)
251	DSL	3/6	544-553	1.1	7.6	Delta serrate ligand
251	Pentaxin	1/1	605-743	80.5	1.6e-25	Pentaxin family
251	Sushi	1/31	766-820	47.3	3.1e-11	Sushi domain (SCR repeat)
251	Sushi	2/31	825-878	68.8	1.4e-17	Sushi domain (SCR repeat)
251	Paramecium_SA	1/4	871-885	5.3	0.19	Paramecium surface antigen domain
251	EGF	7/10	884-918	29.9	1.9e-07	EGF-like domain
251	Sushi	3/31	924-977	62.9	8.7e-16	Sushi domain (SCR repeat)
251	Sushi	4/31	982-1035	58.5	1.8e-14	Sushi domain (SCR repeat)
251	Sushi	5/31	1040-1093	57.5	3.7e-14	Sushi domain (SCR repeat)
251	Sushi	6/31	1098-1151	56.3	8.1e-14	Sushi domain (SCR repeat)
251	Sushi	7/31	1156-1213	30.6	6e-07	Sushi domain (SCR repeat)
251	Sushi	8/31	1218-1276	39.4	3.3e-09	Sushi domain (SCR repeat)
251	Sushi	9/31	1281-1334	71.9	1.7e-18	Sushi domain (SCR repeat)
251	Sushi	10/31	1339-1391	48.3	1.7e-11	Sushi domain (SCR repeat)
251	Sushi	11/31	1399-1453	67.3	4.1e-17	Sushi domain (SCR repeat)
251	Sushi	12/31	1458-1511	38.9	4.3e-09	Sushi domain (SCR repeat)
251	Sushi	13/31	1516-1570	56.3	8.5e-14	Sushi domain (SCR repeat)
251	Sushi	14/31	1575-1628	48.6	1.4e-11	Sushi domain (SCR repeat)

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
251	Paramecium_SA	2/4	1621-1634	0.1	9.7	Paramecium surface antigen domain
251	Sushi	15/31	1633-1686	62.1	1.5e-15	Sushi domain (SCR repeat)
251	Sushi	16/31	1691-1743	53.8	4.7e-13	Sushi domain (SCR repeat)
251	HRM	1/2	1710-1764	8.3	0.12	Hormone receptor domain
251	Sushi	17/31	1748-1760	3.7	4.7	Sushi domain (SCR repeat)
251	Sushi	18/31	1795-1847	51.8	1.9e-12	Sushi domain (SCR repeat)
251	Paramecium_SA	3/4	1839-1853	8.5	0.018	Paramecium surface antigen domain
251	Sushi	19/31	1852-1905	44.0	2.2e-10	Sushi domain (SCR repeat)
251	Sushi	20/31	1910-1963	58.2	2.3e-14	Sushi domain (SCR repeat)
251	Laminin_EGF	7/10	1935-1950	0.5	16	Laminin EGF-like (Domains III and V)
251	TIL	1/2	1935-1950	5.9	0.18	Trypsin Inhibitor like cysteine rich do
251	Sushi	21/31	1968-2021	60.4	4.8e-15	Sushi domain (SCR repeat)
251	Paramecium_SA	4/4	2014-2027	1.2	4.2	Paramecium surface antigen domain
251	Sushi	22/31	2026-2079	51.0	3.3e-12	Sushi domain (SCR repeat)
251	Sushi	23/31	2084-2137	54.3	3.3e-13	Sushi domain (SCR repeat)
251	HRM	2/2	2118-2130	5.6	0.69	Hormone receptor domain
251	Sushi	24/31	2142-2194	38.7	5.1e-09	Sushi domain (SCR repeat)
251	TIL	2/2	2181-2201	1.3	5.3	Trypsin Inhibitor like cysteine rich do
251	Sushi	25/31	2199-2252	48.1	2e-11	Sushi domain (SCR repeat)
251	Sushi	26/31	2257-2311	47.1	3.4e-11	Sushi domain (SCR repeat)
251	Laminin_EGF	8/10	2282-2298	2.6	3.7	Laminin EGF-like (Domains III and V)
251	Sushi	27/31	2316-2365	31.4	3.6e-07	Sushi domain (SCR repeat)
251	Sushi	28/31	2376-2429	53.7	5e-13	Sushi domain (SCR repeat)
251	Sushi	29/31	2434-2487	46.6	4.7e-11	Sushi domain (SCR repeat)
251	Sushi	30/31	2492-2546	42.1	6.7e-10	Sushi domain (SCR repeat)
251	Sushi	31/31	2551-2603	53.3	6.6e-13	Sushi domain (SCR repeat)
251	C_tripleX	1/2	2597-2613	6.8	0.17	Cysteine rich repeat
251	EGF	8/10	2603-2634	22.6	1.9e-05	EGF-like domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
251	Laminin_EGF	9/10	2622-2636	1.6	7.2	Laminin EGF-like (Domains III and V)
251	DSL	4/6	2624-2634	5.6	0.32	Delta serrate ligand
251	EGF	9/10	2639-2666	29.9	1.9e-07	EGF-like domain
251	Laminin_EGF	10/10	2644-2666	4.2	1.3	Laminin EGF-like (Domains III and V)
251	DSL	5/6	2657-2666	6.7	0.15	Delta serrate ligand
251	C_tripleX	2/2	2669-2683	1.3	12	Cysteine rich repeat
251	EGF	10/10	2671-2698	22.5	2.1e-05	EGF-like domain
251	DSL	6/6	2689-2698	2.3	3.2	Delta serrate ligand
252	JmjC	1/1	174-288	141.3	5.2e-41	jmjC domain
255	PSI	1/1	327-372	23.6	5.9e-07	Plexin repeat
256	SNF7	1/1	6-176	163.3	4.9e-46	SNF7
257	DUF323	1/1	87-342	389.0	6e-114	Domain of unknown function (DUF323)
258	Lectin_C	1/1	53-164	127.9	2.2e-35	Lectin C-type domain
259	ARD	1/1	3-157	279.6	5.1e-81	ARD/ARD' family
259	AraC_binding	1/1	85-138	10.6	0.015	AraC-like ligand binding domain
260	Metallophos	1/1	70-285	49.1	1.3e-12	Calcineurin-like phosphoesterase
261	Reprolysin	1/1	218-286	19.3	6e-05	Reprolysin (M12B) family zinc metallopr
261	Peptidase_M43	1/1	224-234	6.3	0.081	Pregnancy-associated plasma protein-A
261	TSP_1	1/7	388-438	48.2	7.4e-14	Thrombospondin type 1 domain
261	TSP_1	2/7	694-704	4.5	0.89	Thrombospondin type 1 domain
261	TSP_1	3/7	735-742	1.4	7.5	Thrombospondin type 1 domain
261	TSP_1	4/7	753-804	4.1	1.2	Thrombospondin type 1 domain
261	TSP_1	5/7	961-1012	5.4	0.5	Thrombospondin type 1 domain
261	TSP_1	6/7	1023-1047	8.5	0.057	Thrombospondin type 1 domain
261	TSP_1	7/7	1079-1102	12.1	0.0049	Thrombospondin type 1 domain
262	IgaA	1/1	226-255	5.0	0.047	Intracellular growth attenuator protein IgaA
263	Herpes_ORF11	1/1	37-79	8.0	0.018	Herpesvirus dUTPase protein
263	ig	1/3	60-133	7.3	0.42	Immunoglobulin domain
263	ig	2/3	171-224	10.6	0.054	Immunoglobulin domain
263	ig	3/3	280-339	1.0	20	Immunoglobulin domain
265	Herpes_ORF11	1/1	37-79	8.0	0.018	Herpesvirus dUTPase protein
265	ig	1/3	60-133	7.3	0.42	Immunoglobulin domain
265	ig	2/3	171-224	10.6	0.054	Immunoglobulin domain
265	ig	3/3	280-339	1.0	20	Immunoglobulin domain
266	Herpes_ORF11	1/1	37-79	8.0	0.018	Herpesvirus dUTPase protein
266	ig	1/3	60-133	7.3	0.42	Immunoglobulin domain
266	ig	2/3	171-224	10.6	0.054	Immunoglobulin domain
266	ig	3/3	280-339	1.0	20	Immunoglobulin domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
267	Herpes_ORF11	1/1	57-93	6.7	0.045	Herpesvirus dUTPase protein
267	ig	1/3	74-147	7.3	0.42	Immunoglobulin domain
267	ig	2/3	185-238	10.6	0.054	Immunoglobulin domain
267	ig	3/3	294-353	1.0	20	Immunoglobulin domain
268	ig	1/1	53-115	25.4	5.8e-06	Immunoglobulin domain
269	NPDC1	1/2	1-23	33.9	1.8e-09	Neural proliferation differentiation control
269	NPDC1	2/2	24-165	401.4	1.1e-117	Neural proliferation differentiation control
270	AdoHcyase	1/2	41-177	209.7	4.7e-63	S-adenosyl-L-homocysteine hydrolase
270	AdoHcyase	2/2	181-468	170.5	2.5e-51	S-adenosyl-L-homocysteine hydrolase
270	AdoHcyase_NAD	1/1	228-389	310.9	1.9e-90	S-adenosyl-L-homocysteine hydrolase, NA
271	ig	1/4	34-117	35.0	1.6e-08	Immunoglobulin domain
271	ig	2/4	164-229	21.3	7.5e-05	Immunoglobulin domain
271	ig	3/4	281-350	6.7	0.6	Immunoglobulin domain
271	ig	4/4	387-454	35.1	1.5e-08	Immunoglobulin domain
272	Ifi-6-16	1/1	16-98	159.7	7.2e-46	Interferon-induced 6-16 family
273	REV	1/2	48-63	3.3	0.87	REV protein (anti-repression trans-act
273	REV	2/2	148-163	3.3	0.87	REV protein (anti-repression trans-act
273	Pox_A_type_inc	1/1	228-250	10.3	0.041	Viral A-type inclusion protein repeat
273	Pentaxin	1/1	342-519	107.1	6e-34	Pentaxin family
275	fn3	1/6	39-102	13.8	0.0021	Fibronectin type III domain
275	VWA	1/1	186-358	223.9	6.1e-70	von Willebrand factor type A domain
275	fn3	2/6	384-467	52.5	1.5e-14	Fibronectin type III domain
275	fn3	3/6	474-552	65.1	3.7e-18	Fibronectin type III domain
275	fn3	4/6	564-646	31.0	2.4e-08	Fibronectin type III domain
275	fn3	5/6	654-734	46.6	7.7e-13	Fibronectin type III domain
275	fn3	6/6	747-827	59.1	1.9e-16	Fibronectin type III domain
275	TSP_N	1/1	849-1044	128.0	6.9e-39	Thrombospondin N-terminal -like domain
275	Collagen	1/3	1079-1122	34.1	2.1e-08	Collagen triple helix repeat (20 copies)
275	Collagen	2/3	1124-1180	52.4	2.9e-13	Collagen triple helix repeat (20 copies)
275	Collagen	3/3	1255-1271	7.4	0.27	Collagen triple helix repeat (20 copies)
276	fn3	1/6	39-102	13.8	0.0021	Fibronectin type III domain
276	VWA	1/1	186-358	223.9	6.1e-70	von Willebrand factor type A domain
276	fn3	2/6	384-467	52.5	1.5e-14	Fibronectin type III domain
276	fn3	3/6	474-552	65.1	3.7e-18	Fibronectin type III domain
276	fn3	4/6	564-646	31.0	2.4e-08	Fibronectin type III domain
276	fn3	5/6	654-734	46.6	7.7e-13	Fibronectin type III domain
276	fn3	6/6	747-827	59.1	1.9e-16	Fibronectin type III domain
276	TSP_N	1/1	849-1044	128.0	6.9e-39	Thrombospondin N-terminal -like domai
276	Collagen	1/4	1078-1132	31.8	8.4e-08	Collagen triple helix repeat (20 copi

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
276	Collagen	2/4	1134-1173	26.9	1.7e-06	Collagen triple helix repeat (20 copi
276	Collagen	3/4	1174-1230	52.4	2.9e-13	Collagen triple helix repeat (20 copi
276	Collagen	4/4	1305-1321	7.4	0.27	Collagen triple helix repeat (20 copi
277	fn3	1/6	39-102	13.8	0.0021	Fibronectin type III domain
277	VWA	1/1	186-358	223.9	6.1e-70	von Willebrand factor type A domain
277	fn3	2/6	384-467	52.5	1.5e-14	Fibronectin type III domain
277	fn3	3/6	474-552	65.1	3.7e-18	Fibronectin type III domain
277	fn3	4/6	564-646	31.0	2.4e-08	Fibronectin type III domain
277	fn3	5/6	654-734	46.6	7.7e-13	Fibronectin type III domain
277	fn3	6/6	747-827	59.1	1.9e-16	Fibronectin type III domain
277	TSP_N	1/1	849-1044	128.0	6.9e-39	Thrombospondin N-terminal -like domain
277	Collagen	1/3	1078-1135	43.2	8e-11	Collagen triple helix repeat (20 copies)
277	Collagen	2/3	1142-1198	52.4	2.9e-13	Collagen triple helix repeat (20 copies)
277	Collagen	3/3	1273-1289	7.4	0.27	Collagen triple helix repeat (20 copies)
278	Nop52	1/2	8-52	73.8	1.5e-19	Nucleolar protein,Nop52
278	Nop52	2/2	53-202	315.0	1.1e-91	Nucleolar protein,Nop52
279	LRR	1/4	65-88	1.7	17	Leucine Rich Repeat
279	LRR	2/4	89-112	11.0	0.04	Leucine Rich Repeat
279	LRR	3/4	113-136	8.6	0.19	Leucine Rich Repeat
279	LRR	4/4	137-160	17.2	0.00071	Leucine Rich Repeat
279	LRRCT	1/1	194-219	17.3	0.00017	Leucine rich repeat C-terminal domain
279	EPTP	1/4	223-263	66.5	4.6e-17	EPTP domain
279	EPTP	2/4	292-309	2.1	13	EPTP domain
279	EPTP	3/4	411-452	85.4	1.4e-22	EPTP domain
279	EPTP	4/4	456-483	9.3	0.14	EPTP domain
281	7tm_1	1/2	86-124	8.2	0.0037	7 transmembrane receptor (rhodopsin family)
281	7tm_1	2/2	315-338	2.1	0.42	7 transmembrane receptor (rhodopsin family)
282	LRRNT	1/2	73-102	29.1	5.4e-08	Leucine rich repeat N-terminal domain
282	LRR	1/21	104-127	10.1	0.072	Leucine Rich Repeat
282	LRR	2/21	128-151	11.9	0.022	Leucine Rich Repeat
282	LRR	3/21	152-175	10.7	0.048	Leucine Rich Repeat
282	LRR	4/21	176-199	12.1	0.02	Leucine Rich Repeat
282	LRR	5/21	200-223	9.3	0.12	Leucine Rich Repeat
282	LRR	6/21	224-247	13.2	0.0095	Leucine Rich Repeat
282	LRR	7/21	248-271	6.0	1	Leucine Rich Repeat
282	LRR	8/21	272-295	0.1	48	Leucine Rich Repeat
282	LRR	9/21	296-319	5.2	1.8	Leucine Rich Repeat
282	LRR	10/21	320-341	13.5	0.0081	Leucine Rich Repeat
282	LRR	11/21	342-392	4.5	2.7	Leucine Rich Repeat
282	LRRCT	1/2	377-399	4.5	1.8	Leucine rich repeat C-terminal domain
282	LRRNT	2/2	436-465	17.0	0.00024	Leucine rich repeat N-terminal domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
282	LRR	12/21	467-490	9.3	0.12	Leucine Rich Repeat
282	LRR	13/21	491-514	7.3	0.43	Leucine Rich Repeat
282	LRR	14/21	515-538	10.7	0.049	Leucine Rich Repeat
282	LRR	15/21	539-562	10.3	0.064	Leucine Rich Repeat
282	LRR	16/21	587-610	9.1	0.14	Leucine Rich Repeat
282	LRR	17/21	611-634	15.5	0.0021	Leucine Rich Repeat
282	LRR	18/21	635-658	6.0	1	Leucine Rich Repeat
282	LRR	19/21	660-683	10.7	0.048	Leucine Rich Repeat
282	LRR	20/21	685-706	10.8	0.047	Leucine Rich Repeat
282	LRR	21/21	707-735	6.1	0.95	Leucine Rich Repeat
282	LRRCT	2/2	739-764	13.0	0.0037	Leucine rich repeat C-terminal domain
283	SCAN	1/1	45-140	190.9	2.4e-54	SCAN domain
283	zf-C2H2	1/8	232-254	4.4	6.6	Zinc finger, C2H2 type
283	XPA_N	1/5	280-292	3.0	4	XPA protein N-terminal
283	TFIIS_C	1/6	283-293	5.0	0.72	Transcription factor S-II (TFIIS)
283	zf-C2H2	2/8	283-305	30.2	2.8e-06	Zinc finger, C2H2 type
283	zf-BED	1/6	284-306	2.0	6	BED zinc finger
283	XPA_N	2/5	308-320	4.6	1.3	XPA protein N-terminal
283	TFIIS_C	2/6	311-321	8.3	0.067	Transcription factor S-II (TFIIS)
283	zf-C2H2	3/8	311-333	33.2	5e-07	Zinc finger, C2H2 type
283	zf-BED	2/6	312-334	3.8	1.8	BED zinc finger
283	TFIIS_C	3/6	339-349	2.3	4.7	Transcription factor S-II (TFIIS)
283	zf-C2H2	4/8	339-361	25.2	4.8e-05	Zinc finger, C2H2 type
283	zf-BED	3/6	341-362	11.5	0.0094	BED zinc finger
283	XPA_N	3/5	364-376	1.7	9	XPA protein N-terminal
283	TFIIS_C	4/6	367-377	5.0	0.72	Transcription factor S-II (TFIIS)
283	zf-C2H2	5/8	367-389	26.7	2e-05	Zinc finger, C2H2 type
283	zf-BED	4/6	381-390	0.7	14	BED zinc finger
283	XPA_N	4/5	392-404	2.0	7.7	XPA protein N-terminal
283	TFIIS_C	5/6	395-405	5.6	0.45	Transcription factor S-II (TFIIS)
283	zf-C2H2	6/8	395-417	29.6	3.9e-06	Zinc finger, C2H2 type
283	zf-BED	5/6	396-418	7.3	0.16	BED zinc finger
283	zf-C2H2	7/8	423-445	29.8	3.4e-06	Zinc finger, C2H2 type
283	zf-BED	6/6	424-446	1.7	7.4	BED zinc finger
283	XPA_N	5/5	448-460	1.9	7.9	XPA protein N-terminal
283	TFIIS_C	6/6	451-461	4.0	1.5	Transcription factor S-II (TFIIS)
283	zf-C2H2	8/8	451-473	32.0	9.6e-07	Zinc finger, C2H2 type
284	Pep_M12B_propep	1/1	82-208	77.1	1.7e-24	Reprolysin family propeptide
284	Reprolysin	1/2	263-293	4.5	0.61	Reprolysin (M12B) family zinc metallo
284	Reprolysin	2/2	325-422	56.9	3.4e-15	Reprolysin (M12B) family zinc metallo
284	TSP_1	1/1	569-591	15.2	0.00056	Thrombospondin type 1 domain
284	ADAM_spacer1	1/1	691-799	169.4	7.2e-48	ADAM-TS Spacer 1
285	Endomucin	1/1	1-261	552.7	3.1e-163	Endomucin
286	Dor1	1/1	32-388	684.1	8.3e-203	Dor1-like family
287	WH2	1/1	727-744	21.2	4.1e-05	WH2 motif
291	SAM	1/1	135-198	42.8	1.3e-11	SAM domain (Sterile alpha motif)
292	E1-E2_ATPase	1/1	126-164	8.6	0.017	E1-E2 ATPase
292	Hydrolase	1/2	401-747	28.4	8.1e-08	haloacid dehalogenase-like hydrolase

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
292	Hydrolase	2/2	816-842	11.2	0.0068	haloacid dehalogenase-like hydrolase
292	BPD_transp_1	1/1	1038-1099	12.5	0.0045	Binding-protein-dependent transport syst
293	C2	1/2	12-64	26.7	2.2e-07	C2 domain
293	C2	2/2	112-195	53.4	2.9e-15	C2 domain
293	Copine	1/1	275-422	338.3	1.1e-98	Copine
294	NIDO	1/1	206-295	126.0	8.2e-36	Nidogen-like
294	VWC	1/3	303-344	10.7	0.015	von Willebrand factor type C domain
294	LRRNT	1/3	321-339	4.6	1.3	Leucine rich repeat N-terminal domain
294	VWD	1/4	365-521	191.6	1.1e-56	von Willebrand factor type D domain
294	C_tripleX	1/4	540-555	4.1	1.4	Cysteine rich repeat
294	TIL	1/3	640-693	56.8	9.8e-18	Trypsin Inhibitor like cysteine rich d
294	EB	1/3	676-690	3.2	1.8	EB module
294	VWC	2/3	695-733	2.6	3.2	von Willebrand factor type C domain
294	TIL_assoc	1/2	708-749	8.7	0.033	TILa domain
294	LRRNT	2/3	713-728	6.5	0.34	Leucine rich repeat N-terminal domain
294	VWD	2/4	756-909	137.9	9.3e-41	von Willebrand factor type D domain
294	TIL	2/3	1027-1079	47.7	7.9e-15	Trypsin Inhibitor like cysteine rich d
294	C_tripleX	2/4	1050-1061	4.9	0.72	Cysteine rich repeat
294	EB	2/3	1060-1073	4.3	0.83	EB module
294	EGF	1/2	1060-1073	0.3	29	EGF-like domain
294	VWD	3/4	1143-1301	157.4	1.6e-46	von Willebrand factor type D domain
294	C_tripleX	3/4	1320-1330	0.4	25	Cysteine rich repeat
294	TIL	3/3	1415-1468	45.8	3.2e-14	Trypsin Inhibitor like cysteine rich d
294	C_tripleX	4/4	1422-1432	2.3	5.6	Cysteine rich repeat
294	VWC	3/3	1470-1507	6.6	0.22	von Willebrand factor type C domain
294	TIL_assoc	2/2	1485-1523	7.2	0.095	TILa domain
294	LRRNT	3/3	1487-1503	0.9	16	Leucine rich repeat N-terminal domain
294	Dickkopf_N	1/1	1506-1516	7.2	0.09	Dickkopf N-terminal cysteine-rich regi
294	VWD	4/4	1530-1682	150.8	1.5e-44	von Willebrand factor type D domain
294	Zona_pellucida	1/1	1848-2102	253.2	4.4e-73	Zona pellucida-like domain
294	EGF	2/2	2131-2164	18.4	0.00028	EGF-like domain
294	EB	3/3	2150-	1.5	6.6	EB module

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
			2164			
295	NIDO	1/1	43-133	127.5	3.1e-36	Nidogen-like
295	EGF	1/16	147-183	28.9	3.5e-07	EGF-like domain
295	Laminin_EGF	1/9	161-183	4.3	1.2	Laminin EGF-like (Domains III and V)
295	DSL	1/9	173-183	3.1	1.9	Delta serrate ligand
295	EGF	2/16	190-221	38.0	1.1e-09	EGF-like domain
295	Laminin_EGF	2/9	209-222	3.3	2.3	Laminin EGF-like (Domains III and V)
295	EGF	3/16	228-259	30.3	1.4e-07	EGF-like domain
295	Laminin_EGF	3/9	235-260	1.5	7.9	Laminin EGF-like (Domains III and V)
295	DSL	2/9	249-259	0.2	14	Delta serrate ligand
295	EGF	4/16	266-297	43.5	3.2e-11	EGF-like domain
295	EGF	5/16	308-339	42.5	5.9e-11	EGF-like domain
295	DSL	3/9	330-339	2.8	2.3	Delta serrate ligand
295	EGF	6/16	349-374	12.7	0.011	EGF-like domain
295	EGF	7/16	383-407	11.3	0.026	EGF-like domain
295	EGF	8/16	420-451	35.3	5.9e-09	EGF-like domain
295	Cripto	1/5	440-461	0.8	6	Cripto growth factor
295	DSL	4/9	442-451	0.2	14	Delta serrate ligand
295	Prokineticin	1/2	457-480	3.8	0.33	Prokineticin
295	EGF	9/16	459-490	30.2	1.5e-07	EGF-like domain
295	Cripto	2/5	464-491	6.4	0.15	Cripto growth factor
295	Laminin_EGF	4/9	478-491	3.9	1.6	Laminin EGF-like (Domains III and V)
295	EGF	10/16	498-529	41.8	9.6e-11	EGF-like domain
295	Prokineticin	2/2	502-519	2.2	1.1	Prokineticin
295	Cripto	3/5	503-530	14.6	0.00068	Cripto growth factor
295	Laminin_EGF	5/9	518-530	9.1	0.05	Laminin EGF-like (Domains III and V)
295	DSL	5/9	520-529	1.1	7.8	Delta serrate ligand
295	EGF	11/16	536-567	31.8	5.4e-08	EGF-like domain
295	Laminin_EGF	6/9	556-567	6.4	0.29	Laminin EGF-like (Domains III and V)
295	DSL	6/9	558-567	2.3	3.3	Delta serrate ligand
295	Sushi	1/1	573-626	28.7	1.8e-06	Sushi domain (SCR repeat)
295	EGF	12/16	632-663	34.5	1e-08	EGF-like domain
295	DSL	7/9	653-663	3.3	1.7	Delta serrate ligand
295	EGF	13/16	670-701	35.7	4.6e-09	EGF-like domain
295	Laminin_EGF	7/9	689-702	3.6	2	Laminin EGF-like (Domains III and V)
295	DSL	8/9	691-701	0.8	9.3	Delta serrate ligand
295	Laminin_EGF	8/9	706-740	1.1	10	Laminin EGF-like (Domains III and V)
295	EGF	14/16	708-739	30.4	1.4e-07	EGF-like domain
295	Cripto	4/5	728-740	1.2	4.6	Cripto growth factor
295	EGF	15/16	746-777	29.0	3.4e-07	EGF-like domain
295	DSL	9/9	767-777	4.9	0.54	Delta serrate ligand
295	fn3	1/3	781-862	38.6	1.6e-10	Fibronectin type III domain
295	fn3	2/3	880-963	42.8	9.4e-12	Fibronectin type III domain
295	fn3	3/3	979-1061	45.7	1.4e-12	Fibronectin type III domain
295	EGF	16/16	1186-1217	38.7	6.8e-10	EGF-like domain
295	Cripto	5/5	1191-	7.7	0.062	Cripto growth factor

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
			1218			
295	Laminin_EGF	9/9	1206-1218	2.5	3.9	Laminin EGF-like (Domains III and V)
297	zf-C3HC4	1/1	325-365	35.8	1.4e-12	Zinc finger, C3HC4 type (RING finger)
297	Prominin	1/1	18-98	121.3	3.3e-34	Prominin
298	Prominin	1/1	1-669	1328.5	0	Prominin
299	Prominin	1/1	18-206	364.0	2e-106	Prominin
301	zf-C2H2	1/3	86-110	27.6	1.2e-05	Zinc finger, C2H2 type
301	zf-C2H2	2/3	116-140	32.6	6.8e-07	Zinc finger, C2H2 type
301	zf-C2H2	3/3	146-168	29.5	4e-06	Zinc finger, C2H2 type
303	ig	1/1	183-237	31.0	2e-07	Immunoglobulin domain
305	Pkinase	1/2	39-219	195.0	1.4e-57	Protein kinase domain
305	DUF244	1/1	284-313	4.9	0.086	Uncharacterized protein family (ORF7) DUF
305	Pkinase	2/2	307-324	8.9	0.013	Protein kinase domain
306	7tm_1	1/1	58-303	265.7	5.6e-89	7 transmembrane receptor (rhodopsin family)
307	Lectin_C	1/1	135-158	10.8	0.041	Lectin C-type domain
308	Lectin_C	1/1	135-158	10.8	0.041	Lectin C-type domain
310	Lectin_C	1/1	135-158	10.8	0.041	Lectin C-type domain
311	Ank	1/5	48-80	40.0	3.3e-10	Ankyrin repeat
311	Ank	2/5	111-143	34.3	1.3e-08	Ankyrin repeat
311	Ank	3/5	144-166	16.0	0.0016	Ankyrin repeat
311	Ank	4/5	185-217	47.1	3.4e-12	Ankyrin repeat
311	Ank	5/5	220-249	26.8	1.6e-06	Ankyrin repeat
311	SH3	1/1	298-337	14.0	0.0049	SH3 domain
311	SAM	1/2	492-555	73.6	2e-20	SAM domain (Sterile alpha motif)
311	SAM	2/2	726-780	57.1	1e-15	SAM domain (Sterile alpha motif)
312	LRRNT	1/1	23-49	15.2	0.0008	Leucine rich repeat N-terminal domain
312	LRR	1/5	51-74	18.3	0.00034	Leucine Rich Repeat
312	LRR	2/5	75-98	13.0	0.011	Leucine Rich Repeat
312	LRR	3/5	99-122	10.4	0.058	Leucine Rich Repeat
312	LRR	4/5	123-146	18.3	0.00034	Leucine Rich Repeat
312	LRR	5/5	147-175	1.3	22	Leucine Rich Repeat
312	LRRCT	1/1	183-208	19.1	4.5e-05	Leucine rich repeat C-terminal domain
312	ig	1/4	224-283	35.1	1.6e-08	Immunoglobulin domain
312	ig	2/4	320-376	37.1	4.5e-09	Immunoglobulin domain
312	ig	3/4	416-466	22.3	4e-05	Immunoglobulin domain
312	ig	4/4	501-558	33.7	3.7e-08	Immunoglobulin domain
312	An_peroxidase	1/1	701-1251	649.9	1.6e-192	Animal haem peroxidase
312	IFP_35_N	1/1	1344-1366	9.1	0.029	Interferon-induced 35 kDa protein (IFP
312	TIL_assoc	1/1	1370-1409	16.9	0.0001	TILa domain
312	VWC	1/1	1371-1426	38.0	2.3e-10	von Willebrand factor type C domain
313	TPR	1/2	82-115	27.7	6.3e-07	TPR Domain
313	TPR	2/2	116-138	11.9	0.02	TPR Domain
313	zf-CCCH	1/4	494-503	8.3	0.13	Zinc finger C-x8-C-x5-C-x3-H type (and simil
313	zf-CCCH	2/4	625-637	8.9	0.08	Zinc finger C-x8-C-x5-C-x3-H

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
						type (and simil
313	zf-CCCH	3/4	755-781	18.0	0.00015	Zinc finger C-x8-C-x5-C-x3-H type (and simil
313	zf-C2H2	1/1	842-866	14.9	0.017	Zinc finger, C2H2 type
313	zf-CCCH	4/4	887-913	22.8	5.3e-06	Zinc finger C-x8-C-x5-C-x3-H type (and simil
314	Torsin	1/2	106-123	18.9	2.2e-06	Torsin
314	Torsin	2/2	124-349	520.3	9.7e-172	Torsin
315	ig	1/3	71-150	23.5	1.9e-05	Immunoglobulin domain
315	ig	2/3	186-248	2.7	7.1	Immunoglobulin domain
315	ig	3/3	284-340	14.6	0.0047	Immunoglobulin domain
316	EBP	1/1	23-222	454.9	8.6e-134	Emopamil binding protein
318	FG-GAP	1/5	46-88	21.4	2.5e-05	FG-GAP repeat
318	FG-GAP	2/5	105-147	21.9	1.9e-05	FG-GAP repeat
318	LacI	1/1	215-230	10.4	0.04	Bacterial regulatory proteins, lacI family
318	FG-GAP	3/5	283-333	23.9	5e-06	FG-GAP repeat
318	FG-GAP	4/5	336-381	0.9	16	FG-GAP repeat
318	FG-GAP	5/5	395-437	21.0	3.4e-05	FG-GAP repeat
319	IRF	1/1	1-76	201.3	2.3e-58	Interferon regulatory factor transcrip
319	Heme_oxygenase	1/1	29-77	9.9	0.013	Heme oxygenase
320	ART	1/1	56-306	192.5	2.3e-55	NAD:arginine ADP-ribosyltransferase
321	C1q	1/1	998-1123	101.5	2e-27	C1q domain
322	Ank	1/1	16-48	33.4	2.3e-08	Ankyrin repeat
322	Clip	1/1	74-118	23.2	7e-06	Clip-like
323	PRA1	1/1	1-156	211.2	1.4e-61	PRA1 family protein
326	Thioredoxin	1/1	3-64	34.1	9e-09	Thioredoxin
326	Evr1_Alr	1/1	349-436	62.5	2.6e-18	Erv1 / Alr family
327	Mito_carr	1/3	9-104	116.1	9.9e-34	Mitochondrial carrier protein
327	Mito_carr	2/3	109-201	120.7	4.5e-35	Mitochondrial carrier protein
327	Mito_carr	3/3	208-298	100.5	4e-29	Mitochondrial carrier protein
328	EF1_GNE	1/1	176-262	187.5	3e-54	EF-1 guanine nucleotide exchange domain
331	Lipocalin	1/1	38-183	60.6	5.7e-16	Lipocalin / cytosolic fatty-acid binding pr
333	Cytochrom_C	1/1	5-103	124.9	1.8e-34	Cytochrome c
334	UPF0191	1/2	256-285	4.3	0.37	Uncharacterised protein family (UPF0191
334	UPF0191	2/2	297-318	11.4	0.0029	Uncharacterised protein family (UPF0191
336	ig	1/5	38-115	26.9	2.4e-06	Immunoglobulin domain
336	ig	2/5	154-210	45.3	2.9e-11	Immunoglobulin domain
336	ig	3/5	243-305	32.4	8.1e-08	Immunoglobulin domain
336	ig	4/5	339-399	17.3	0.00086	Immunoglobulin domain
336	ig	5/5	435-490	25.9	4.5e-06	Immunoglobulin domain
336	fn3	1/2	510-598	19.8	4.1e-05	Fibronectin type III domain
336	fn3	2/2	619-702	20.0	3.5e-05	Fibronectin type III domain
337	DUF803	1/1	29-330	581.2	8.3e-172	Protein of unknown function (DUF803)
338	Prefoldin	1/3	5-44	7.6	0.14	Prefoldin subunit
338	Prefoldin	2/3	54-82	0.3	17	Prefoldin subunit

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
338	Spectrin	1/7	59-121	15.0	0.00079	Spectrin repeat
338	Spectrin	2/7	124-226	22.2	6.3e-06	Spectrin repeat
338	Spectrin	3/7	229-340	25.7	6.1e-07	Spectrin repeat
338	GSPII_E_N	1/1	265-290	7.7	0.082	GSPII_E N-terminal domain
338	Spectrin	4/7	343-449	19.8	3.3e-05	Spectrin repeat
338	Spectrin	5/7	452-538	23.7	2.4e-06	Spectrin repeat
338	Spectrin	6/7	758-865	47.2	3.6e-13	Spectrin repeat
338	Spectrin	7/7	915-976	1.3	7.4	Spectrin repeat
338	Prefoldin	3/3	948-980	0.7	13	Prefoldin subunit
339	Lung_7-TM_R	1/1	215-435	328.4	9.9e-96	Lung seven transmembrane receptor
340	HlyIII	1/1	140-363	230.7	2.7e-66	Haemolysin-III related
340	Glycos_transf_N	1/1	328-346	7.6	0.038	3-Deoxy-D-manno-octulosonic-acid tran
341	Pep_M12B_propep	1/1	33-148	174.5	8.4e-56	Reprolysin family propeptide
341	Reprolysin	1/1	158-355	342.1	7.6e-100	Reprolysin (M12B) family zinc metallo
341	Disintegrin	1/1	373-451	114.5	1.3e-35	Disintegrin
341	EGF	1/2	457-476	2.5	7.5	EGF-like domain
341	EGF	2/2	593-617	11.5	0.024	EGF-like domain
341	SBP56	1/1	606-615	5.8	0.057	56kDa selenium binding protein (SBP56
342	IQ	1/3	447-465	2.6	10	IQ calmodulin-binding motif
342	IQ	2/3	470-490	22.1	1.6e-05	IQ calmodulin-binding motif
342	IQ	3/3	529-549	21.8	1.9e-05	IQ calmodulin-binding motif
343	Collagen	1/4	2-30	18.9	0.00023	Collagen triple helix repeat (20 copies)
343	Collagen	2/4	68-123	28.2	7.8e-07	Collagen triple helix repeat (20 copies)
343	Collagen	3/4	126-146	15.4	0.002	Collagen triple helix repeat (20 copies)
343	Collagen	4/4	148-177	19.1	0.00021	Collagen triple helix repeat (20 copies)
344	ig	1/1	221-351	9.9	0.083	Immunoglobulin domain
344	Pkinase	1/2	549-649	66.9	1e-19	Protein kinase domain
344	Pkinase	2/2	723-884	194.9	1.6e-57	Protein kinase domain
345	SCF	1/1	1-283	704.4	3.2e-211	Stem cell factor
345	FH2	1/1	145-162	8.8	0.032	Formin Homology 2 Domain
346	NACHT	1/1	1-156	210.0	1.4e-61	NACHT domain
347	PAAD_DAPIN	1/1	18-103	41.6	2.2e-11	PAAD/DAPIN/Pyrin domain
347	RNA_helicase	1/1	195-215	7.9	0.036	RNA helicase
347	NACHT	1/1	196-365	252.4	4.4e-74	NACHT domain
348	Fibrinogen_C	1/1	240-457	311.1	5.2e-91	Fibrinogen beta and gamma chains, C-term
349	Fibrinogen_C	1/1	240-457	315.7	2.4e-92	Fibrinogen beta and gamma chains, C-term
350	LBP_BPI_CETP	1/1	22-184	143.1	2.9e-41	LBP / BPI / CETP family, N-terminal do
350	LBP_BPI_CETP_C	1/1	290-454	46.8	7.2e-13	LBP / BPI / CETP family, C-terminal do
351	Oxysterol_BP	1/2	19-270	299.0	7.2e-87	Oxysterol-binding protein
351	Oxysterol_B	2/2	329-429	45.7	5.9e-13	Oxysterol-binding protein

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
	P					
352	APC10	1/1	125-152	10.8	0.0056	Anaphase-promoting complex, subunit 10 (
352	BK_channel_a	1/1	1069-1082	4.3	0.079	Calcium-activated BK potassium channel a
352	ZZ	1/2	1598-1641	26.4	1.4e-06	Zinc finger, ZZ type
352	ZZ	2/2	1642-1686	32.1	3.7e-08	Zinc finger, ZZ type
352	NifQ	1/1	1652-1673	7.3	0.058	NifQ
354	Collagen	1/2	37-64	18.8	0.00025	Collagen triple helix repeat (20 copies)
354	Collagen	2/2	65-124	48.8	2.6e-12	Collagen triple helix repeat (20 copies)
354	C1q	1/1	134-258	148.4	1.6e-41	C1q domain
355	Ion_trans	1/2	70-192	29.3	6.3e-08	Ion transport protein
355	Ion_trans	2/2	237-318	7.2	0.094	Ion transport protein
356	Ion_trans	1/2	75-197	29.3	6.3e-08	Ion transport protein
356	Ion_trans	2/2	242-323	7.2	0.094	Ion transport protein
357	A2M_N	1/1	6-613	316.7	3.5e-92	Alpha-2-macroglobulin family N-terminal regi
357	A2M	1/1	721-1448	711.7	4.2e-211	Alpha-2-macroglobulin family
358	PAX	1/1	4-142	279.7	4.6e-81	'Paired box' domain
358	Homeobox	1/1	225-281	112.7	8.8e-31	Homeobox domain
359	Collagen	1/1	41-88	37.2	3.1e-09	Collagen triple helix repeat (20 copies)
359	Lectin_C	1/1	135-238	78.4	1.9e-20	Lectin C-type domain
360	Collagen	1/3	24-82	48.3	3.6e-12	Collagen triple helix repeat (20 copie
360	Collagen	2/3	95-154	42.8	1e-10	Collagen triple helix repeat (20 copie
360	Collagen	3/3	155-191	33.6	2.9e-08	Collagen triple helix repeat (20 copie
360	C1q	1/1	203-329	150.7	3.3e-42	C1q domain
361	Keratin_B2	1/1	74-153	26.1	2.4e-07	Keratin, high sulfur B2 protein
362	Keratin_B2	1/1	111-171	27.0	1.3e-07	Keratin, high sulfur B2 protein
363	Xlink	1/1	26-52	10.9	2.3e-05	Extracellular link domain
363	Lectin_C	1/1	34-160	70.5	4.5e-18	Lectin C-type domain
365	Torsin	1/1	106-396	692.3	1.8e-228	Torsin
366	Torsin	1/2	106-123	18.9	2.2e-06	Torsin
366	Torsin	2/2	124-349	520.3	9.7e-172	Torsin
368	CorA	1/1	150-187	7.3	0.014	CorA-like Mg2+ transporter protein
369	Collagen	1/1	61-109	34.2	2e-08	Collagen triple helix repeat (20 copies)
369	C1q	1/1	128-252	117.4	3.3e-32	C1q domain
371	ig	1/1	42-98	17.2	0.00093	Immunoglobulin domain
371	MARVEL	1/1	95-161	8.5	0.036	Membrane-associating domain
373	Bradykinin	1/1	19-29	9.5	0.074	Bradykinin
374	SH2	1/2	10-87	103.2	7.2e-35	SH2 domain
374	SH2	2/2	163-239	107.3	2.8e-36	SH2 domain
374	Pkinase	1/1	338-593	264.5	4.5e-78	Protein kinase domain
375	SCP	1/1	66-205	167.3	4.1e-49	SCP-like extracellular protein
375	LCCL	1/2	293-384	145.3	2.8e-42	LCCL domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
375	LCCL	2/2	394-492	172.4	2.9e-50	LCCL domain
379	CD20	1/1	24-56	11.3	0.0059	CD20/IgE Fc receptor beta subunit family
381	Radical_SAM	1/1	131-296	96.6	1.5e-26	Radical SAM superfamily
382	Dak2	1/1	28-44	9.1	0.035	DAK2 domain
383	Hemopexin	1/3	20-30	0.6	20	Hemopexin
383	Peptidase_M10	1/2	23-69	100.7	3.7e-27	Matrixin
383	Peptidase_M10_N	1/1	79-120	88.6	1.5e-31	Matrix metalloprotease, N-terminal do
383	Peptidase_M10	2/2	127-231	189.0	9.4e-54	Matrixin
383	Fragilysin	1/1	238-263	9.8	0.0052	Fragilysin metallopeptidase (M10C) en
383	Hemopexin	2/3	309-350	46.8	3.2e-13	Hemopexin
383	Hemopexin	3/3	352-366	2.9	4.2	Hemopexin
384	Collagen	1/10	2-58	42.7	1.1e-10	Collagen triple helix repeat (20 copi
384	Collagen	2/10	59-118	50.8	7.6e-13	Collagen triple helix repeat (20 copi
384	Collagen	3/10	122-181	51.9	3.9e-13	Collagen triple helix repeat (20 copi
384	Collagen	4/10	182-241	40.6	3.8e-10	Collagen triple helix repeat (20 copi
384	Collagen	5/10	242-301	51.8	4e-13	Collagen triple helix repeat (20 copi
384	Collagen	6/10	303-350	40.4	4.5e-10	Collagen triple helix repeat (20 copi
384	Collagen	7/10	351-406	40.5	4.2e-10	Collagen triple helix repeat (20 copi
384	Collagen	8/10	408-462	40.4	4.3e-10	Collagen triple helix repeat (20 copi
384	Collagen	9/10	465-524	38.9	1.1e-09	Collagen triple helix repeat (20 copi
384	Collagen	10/10	525-584	42.8	1e-10	Collagen triple helix repeat (20 copi
384	COLFI	1/2	639-697	92.7	6.9e-38	Fibrillar collagen C-terminal domain
384	COLFI	2/2	706-822	56.8	2.1e-23	Fibrillar collagen C-terminal domain
387	DUF28	1/1	61-297	189.1	1.7e-55	Domain of unknown function DUF28
392	7tm_1	1/1	68-322	159.7	1.1e-53	7 transmembrane receptor (rhodopsin fa
392	Spore_permease	1/1	251-281	9.0	0.021	Spore germination protein
393	7tm_1	1/1	51-305	159.7	1.1e-53	7 transmembrane receptor (rhodopsin fa
393	Spore_permease	1/1	234-264	9.0	0.021	Spore germination protein
395	FCH	1/1	14-102	78.9	4.5e-21	Fes/CIP4 homology domain
395	SH3	1/1	366-422	70.3	2.5e-18	SH3 domain
396	HSP70	1/1	3-380	364.0	1.9e-106	Hsp70 protein
397	ig	1/5	54-112	2.3	8.7	Immunoglobulin domain
397	ig	2/5	150-207	25.0	7.6e-06	Immunoglobulin domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
397	ig	3/5	242-291	28.5	9e-07	Immunoglobulin domain
397	ig	4/5	367-385	12.5	0.017	Immunoglobulin domain
397	ig	5/5	420-439	7.1	0.47	Immunoglobulin domain
398	ig	1/3	38-111	5.5	1.2	Immunoglobulin domain
398	ig	2/3	149-206	25.0	7.6e-06	Immunoglobulin domain
398	ig	3/3	241-290	28.5	9e-07	Immunoglobulin domain
399	ig	1/3	159-217	2.3	8.7	Immunoglobulin domain
399	ig	2/3	255-312	25.0	7.6e-06	Immunoglobulin domain
399	ig	3/3	347-396	28.5	9e-07	Immunoglobulin domain
400	Pep_M12B_propep	1/1	75-191	106.1	8.3e-34	Reprolysin family propeptide
400	Reprolysin	1/1	341-370	22.9	6.2e-06	Reprolysin (M12B) family zinc metallo
400	Disintegrin	1/1	419-494	106.4	4.6e-33	Disintegrin
401	Pep_M12B_propep	1/1	75-191	104.6	2.5e-33	Reprolysin family propeptide
402	Serpin	1/1	43-415	745.5	2.2e-223	Serpin (serine protease inhibitor)
403	KRAB	1/1	39-79	89.1	5.6e-24	KRAB box
403	XPA_N	1/7	201-213	2.4	5.7	XPA protein N-terminal
403	TFIIS_C	1/10	204-214	6.3	0.27	Transcription factor S-II (TFIIS)
403	zf-C2H2	1/16	204-223	27.2	1.5e-05	Zinc finger, C2H2 type
403	zf-BED	1/6	206-223	5.1	0.71	BED zinc finger
403	TFIIS_C	2/10	232-242	2.0	6	Transcription factor S-II (TFIIS)
403	zf-C2H2	2/16	232-254	30.5	2.3e-06	Zinc finger, C2H2 type
403	zf-C2H2	3/16	260-282	24.3	7.7e-05	Zinc finger, C2H2 type
403	zf-C2H2	4/16	288-310	27.4	1.4e-05	Zinc finger, C2H2 type
403	zf-C2H2	5/16	316-338	17.0	0.0051	Zinc finger, C2H2 type
403	XPA_N	2/7	341-353	1.2	12	XPA protein N-terminal
403	TFIIS_C	3/10	344-354	2.8	3.4	Transcription factor S-II (TFIIS)
403	zf-C2H2	6/16	344-366	28.2	8.3e-06	Zinc finger, C2H2 type
403	zf-BED	2/6	345-367	3.3	2.5	BED zinc finger
403	TFIIS_C	4/10	372-382	1.4	9.4	Transcription factor S-II (TFIIS)
403	zf-C2H2	7/16	372-394	18.1	0.0027	Zinc finger, C2H2 type
403	zf-C2H2	8/16	400-422	25.9	3.1e-05	Zinc finger, C2H2 type
403	zf-BED	3/6	401-423	9.7	0.031	BED zinc finger
403	TFIIS_C	5/10	428-438	2.6	4	Transcription factor S-II (TFIIS)
403	zf-C2H2	9/16	428-450	29.7	3.5e-06	Zinc finger, C2H2 type
403	XPA_N	3/7	453-465	2.4	5.9	XPA protein N-terminal
403	TFIIS_C	6/10	456-466	6.4	0.25	Transcription factor S-II (TFIIS)
403	zf-C2H2	10/16	456-478	33.8	3.5e-07	Zinc finger, C2H2 type
403	zf-BED	4/6	457-479	0.0	22	BED zinc finger
403	XPA_N	4/7	481-494	0.6	19	XPA protein N-terminal
403	zf-C2H2	11/16	484-505	19.2	0.0014	Zinc finger, C2H2 type
403	XPA_N	5/7	508-520	4.5	1.4	XPA protein N-terminal
403	TFIIS_C	7/10	511-521	6.8	0.2	Transcription factor S-II (TFIIS)
403	zf-C2H2	12/16	511-533	25.4	4.1e-05	Zinc finger, C2H2 type
403	TFIIS_C	8/10	539-549	2.2	5.3	Transcription factor S-II (TFIIS)
403	zf-C2H2	13/16	539-561	34.3	2.6e-07	Zinc finger, C2H2 type
403	zf-C2H2	14/16	567-589	24.8	5.8e-05	Zinc finger, C2H2 type
403	XPA_N	6/7	592-604	2.4	5.6	XPA protein N-terminal
403	TFIIS_C	9/10	595-605	4.8	0.82	Transcription factor S-II (TFIIS)
403	zf-C2H2	15/16	595-617	21.4	0.0004	Zinc finger, C2H2 type
403	zf-BED	5/6	596-608	5.1	0.72	BED zinc finger
403	XPA_N	7/7	620-632	5.0	1.1	XPA protein N-terminal
403	TFIIS_C	10/10	623-633	6.8	0.2	Transcription factor S-II (TFIIS)

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
403	zf-C2H2	16/16	623-645	34.5	2.3e-07	Zinc finger, C2H2 type
403	zf-BED	6/6	624-646	0.6	16	BED zinc finger
404	CLP_protea se	1/2	67-106	75.7	1.2e-21	Clp protease
404	CLP_protea se	2/2	107-197	189.2	1.9e-54	Clp protease
408	zf-C2H2	1/1	174-196	18.9	0.0017	Zinc finger, C2H2 type
410	F-box	1/1	131-171	13.3	0.0064	F-box domain
410	LRR	1/6	251-280	2.9	7.5	Leucine Rich Repeat
410	LRR	2/6	353-378	5.3	1.7	Leucine Rich Repeat
410	LRR	3/6	379-393	9.9	0.082	Leucine Rich Repeat
410	LRR	4/6	405-429	8.3	0.24	Leucine Rich Repeat
410	LRR	5/6	430-454	9.9	0.083	Leucine Rich Repeat
410	LRR	6/6	550-575	1.3	22	Leucine Rich Repeat
411	Collagen	1/3	2-19	10.0	0.055	Collagen triple helix repeat (20 copies)
411	Collagen	2/3	36-84	39.1	9.8e-10	Collagen triple helix repeat (20 copies)
411	Collagen	3/3	87-146	50.3	1e-12	Collagen triple helix repeat (20 copies)
412	EGF	1/8	129-165	22.8	1.8e-05	EGF-like domain
412	EGF	2/8	169-204	21.9	3e-05	EGF-like domain
412	TIL	1/4	187-209	2.4	2.3	Trypsin Inhibitor like cysteine rich do
412	EGF	3/8	238-273	29.9	1.9e-07	EGF-like domain
412	TIL	2/4	257-279	5.4	0.26	Trypsin Inhibitor like cysteine rich do
412	EGF	4/8	279-314	26.1	2.1e-06	EGF-like domain
412	TIL	3/4	299-320	1.0	6.4	Trypsin Inhibitor like cysteine rich do
412	EGF	5/8	320-353	14.1	0.0044	EGF-like domain
412	EGF	6/8	372-407	30.4	1.4e-07	EGF-like domain
412	TIL	4/4	392-413	10.5	0.0062	Trypsin Inhibitor like cysteine rich do
412	TNFR_c6	1/3	655-672	12.1	0.0087	TNFR/NGFR cysteine-rich region
412	TNFR_c6	2/3	759-780	9.6	0.049	TNFR/NGFR cysteine-rich region
412	EGF	7/8	814-828	3.7	3.5	EGF-like domain
412	TNFR_c6	3/3	815-836	2.5	6	TNFR/NGFR cysteine-rich region
412	EGF	8/8	830-845	2.3	8.1	EGF-like domain
412	CUB	1/2	870-908	52.5	1.7e-15	CUB domain
412	CUB	2/2	947-979	18.4	3.1e-05	CUB domain
413	EGF	1/8	3-39	22.8	1.8e-05	EGF-like domain
413	EGF	2/8	43-78	21.9	3e-05	EGF-like domain
413	TIL	1/4	61-83	2.4	2.3	Trypsin Inhibitor like cysteine rich do
413	EGF	3/8	112-147	29.9	1.9e-07	EGF-like domain
413	TIL	2/4	131-153	5.4	0.26	Trypsin Inhibitor like cysteine rich do
413	EGF	4/8	153-188	26.1	2.1e-06	EGF-like domain
413	TIL	3/4	173-194	1.0	6.4	Trypsin Inhibitor like cysteine rich do
413	EGF	5/8	194-227	14.1	0.0044	EGF-like domain
413	EGF	6/8	246-281	30.4	1.4e-07	EGF-like domain
413	TIL	4/4	266-287	10.5	0.0062	Trypsin Inhibitor like cysteine rich do

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
413	TNFR_c6	1/3	529-546	12.1	0.0087	TNFR/NGFR cysteine-rich region
413	TNFR_c6	2/3	633-654	9.6	0.049	TNFR/NGFR cysteine-rich region
413	EGF	7/8	688-702	3.7	3.5	EGF-like domain
413	TNFR_c6	3/3	689-710	2.5	6	TNFR/NGFR cysteine-rich region
413	EGF	8/8	704-719	2.3	8.1	EGF-like domain
413	CUB	1/2	744-782	52.5	1.7e-15	CUB domain
413	CUB	2/2	821-853	18.4	3.1e-05	CUB domain
414	COX6C	1/1	3-75	136.9	3e-41	Cytochrome c oxidase subunit VIc
415	ig	1/2	39-97	16.8	0.0012	Immunoglobulin domain
415	ig	2/2	128-189	46.5	1.4e-11	Immunoglobulin domain
417	ig	1/3	73-80	0.6	25	Immunoglobulin domain
417	ig	2/3	116-123	0.1	34	Immunoglobulin domain
417	ig	3/3	153-206	17.3	0.00087	Immunoglobulin domain
419	ig	1/3	101-120	7.9	0.28	Immunoglobulin domain
419	ig	2/3	161-218	3.7	3.7	Immunoglobulin domain
419	ig	3/3	253-302	30.5	2.6e-07	Immunoglobulin domain
421	UPAR_LY6	1/2	63-88	8.1	0.63	u-PAR/Ly-6 domain
421	UPAR_LY6	2/2	124-138	12.5	0.065	u-PAR/Ly-6 domain
423	SCP	1/1	52-181	125.4	9.1e-37	SCP-like extracellular protein
423	EGF	1/2	225-260	16.6	0.00092	EGF-like domain
423	EGF	2/2	279-291	7.0	0.42	EGF-like domain
424	ig	1/1	55-144	27.3	1.8e-06	Immunoglobulin domain
425	7tm_1	1/1	2-219	85.7	5.3e-29	7 transmembrane receptor (rhodopsin family)
429	SAP_155	1/2	211-236	3.9	1.5	Splicing factor 3B subunit 1 (Spliceos
429	SAP_155	2/2	467-480	5.5	0.57	Splicing factor 3B subunit 1 (Spliceos
432	UPAR_LY6	1/1	23-96	33.6	5.5e-07	u-PAR/Ly-6 domain
432	Toxin_1	1/1	82-96	10.9	0.074	Snake toxin
435	Peptidase_C 54	1/2	109-168	120.4	4.1e-33	Peptidase family C54
435	Peptidase_C 54	2/2	210-407	265.8	6.9e-77	Peptidase family C54
436	ig	1/4	102-121	8.5	0.2	Immunoglobulin domain
436	ig	2/4	162-219	7.9	0.28	Immunoglobulin domain
436	ig	3/4	255-312	9.6	0.099	Immunoglobulin domain
436	ig	4/4	347-396	31.7	1.2e-07	Immunoglobulin domain
437	ig	1/3	102-121	8.5	0.2	Immunoglobulin domain
437	ig	2/3	162-219	12.3	0.019	Immunoglobulin domain
437	ig	3/3	254-303	29.3	5.5e-07	Immunoglobulin domain
438	ig	1/3	107-143	8.8	0.16	Immunoglobulin domain
438	ig	2/3	184-241	4.8	1.9	Immunoglobulin domain
438	ig	3/3	277-364	13.7	0.0082	Immunoglobulin domain
439	TSP_1	1/3	37-81	25.9	3.5e-07	Thrombospondin type 1 domain
439	TSP_1	2/3	308-318	5.5	0.46	Thrombospondin type 1 domain
439	TSP_1	3/3	363-387	17.4	0.00013	Thrombospondin type 1 domain
440	TSP_1	1/6	37-81	25.9	3.5e-07	Thrombospondin type 1 domain
440	TSP_1	2/6	308-318	5.5	0.46	Thrombospondin type 1 domain
440	TSP_1	3/6	380-404	17.4	0.00013	Thrombospondin type 1 domain
440	TSP_1	4/6	444-463	21.1	9.9e-06	Thrombospondin type 1 domain
440	TSP_1	5/6	531-550	19.8	2.3e-05	Thrombospondin type 1 domain
440	TSP_1	6/6	671-683	0.2	17	Thrombospondin type 1 domain
441	TSP_1	1/6	85-129	25.9	3.5e-07	Thrombospondin type 1 domain
441	TSP_1	2/6	356-366	5.5	0.46	Thrombospondin type 1 domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
441	TSP_1	3/6	428-452	17.4	0.00013	Thrombospondin type 1 domain
441	TSP_1	4/6	492-511	21.1	9.9e-06	Thrombospondin type 1 domain
441	TSP_1	5/6	579-598	19.8	2.3e-05	Thrombospondin type 1 domain
441	TSP_1	6/6	719-731	0.2	17	Thrombospondin type 1 domain
442	UPAR_LY6	1/1	23-101	33.3	7e-07	u-PAR/Ly-6 domain
443	UPAR_LY6	1/1	21-94	87.3	3.9e-23	u-PAR/Ly-6 domain
443	Activin_rec p	1/1	86-100	7.5	0.054	Activin types I and II receptor domain
444	UPAR_LY6	1/1	21-55	34.9	2.3e-07	u-PAR/Ly-6 domain
446	LRRNT	1/1	33-60	30.5	2.1e-08	Leucine rich repeat N-terminal domain
446	LRR	1/10	66-85	1.3	21	Leucine Rich Repeat
446	LRR	2/10	86-109	15.7	0.0019	Leucine Rich Repeat
446	LRR	3/10	110-133	9.3	0.12	Leucine Rich Repeat
446	LRR	4/10	134-157	17.6	0.00054	Leucine Rich Repeat
446	LRR	5/10	158-181	12.8	0.013	Leucine Rich Repeat
446	LRR	6/10	182-205	11.0	0.041	Leucine Rich Repeat
446	LRR	7/10	206-229	11.6	0.027	Leucine Rich Repeat
446	LRR	8/10	230-251	5.9	1.1	Leucine Rich Repeat
446	LRR	9/10	254-277	9.6	0.096	Leucine Rich Repeat
446	LRR	10/10	279-302	11.9	0.022	Leucine Rich Repeat
446	LRRCT	1/1	337-362	9.2	0.061	Leucine rich repeat C-terminal domain
447	ig	1/2	159-217	25.2	6.6e-06	Immunoglobulin domain
447	ig	2/2	267-321	24.4	1.1e-05	Immunoglobulin domain
448	Collagen	1/17	1-55	45.4	2e-11	Collagen triple helix repeat (20 copi
448	Collagen	2/17	56-115	75.7	1.2e-19	Collagen triple helix repeat (20 copi
448	Collagen	3/17	116-175	64.9	1.3e-16	Collagen triple helix repeat (20 copi
448	Collagen	4/17	176-235	61.6	9.9e-16	Collagen triple helix repeat (20 copi
448	Collagen	5/17	236-295	61.1	1.3e-15	Collagen triple helix repeat (20 copi
448	Collagen	6/17	296-355	63.9	2.4e-16	Collagen triple helix repeat (20 copi
448	Collagen	7/17	356-415	64.6	1.6e-16	Collagen triple helix repeat (20 copi
448	Collagen	8/17	416-475	62.1	7.3e-16	Collagen triple helix repeat (20 copi
448	Collagen	9/17	476-535	60.6	1.8e-15	Collagen triple helix repeat (20 copi
448	Collagen	10/17	536-595	70.2	5.2e-18	Collagen triple helix repeat (20 copi
448	Collagen	11/17	599-658	68.4	1.6e-17	Collagen triple helix repeat (20 copi
448	Collagen	12/17	659-718	60.5	2e-15	Collagen triple helix repeat (20 copi
448	Collagen	13/17	719-778	59.2	4.4e-15	Collagen triple helix repeat (20 copi
448	Collagen	14/17	779-838	62.7	5.3e-16	Collagen triple helix repeat (20 copi
448	Collagen	15/17	839-898	60.1	2.6e-15	Collagen triple helix repeat (20 copi
448	Collagen	16/17	899-958	74.1	3.7e-19	Collagen triple helix repeat (20

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
						copi
448	Collagen	17/17	959-1012	40.5	4.1e-10	Collagen triple helix repeat (20 copi
448	COLFI	1/1	1065-1283	565.3	1.4e-228	Fibrillar collagen C-terminal domain
449	IL1	1/2	14-34	2.2	2	Interleukin-1 / 18
449	IL1	2/2	62-154	73.6	2.8e-21	Interleukin-1 / 18
450	Trypsin	1/1	56-101	69.9	1.8e-22	Trypsin
451	Trypsin	1/1	28-262	252.6	4.9e-81	Trypsin
452	Arm	1/2	106-122	2.1	11	Armadillo/beta-catenin-like repeat
452	Arm	2/2	299-340	9.5	0.094	Armadillo/beta-catenin-like repeat
453	Collagen	1/11	77-101	14.9	0.0027	Collagen triple helix repeat (20 copi
453	Collagen	2/11	103-118	7.6	0.24	Collagen triple helix repeat (20 copi
453	Collagen	3/11	126-168	34.9	1.3e-08	Collagen triple helix repeat (20 copi
453	Collagen	4/11	173-209	29.3	3.9e-07	Collagen triple helix repeat (20 copi
453	Collagen	5/11	211-235	8.3	0.15	Collagen triple helix repeat (20 copi
453	Collagen	6/11	237-280	32.2	6.7e-08	Collagen triple helix repeat (20 copi
453	Collagen	7/11	281-314	22.7	2.3e-05	Collagen triple helix repeat (20 copi
453	Collagen	8/11	316-375	45.9	1.5e-11	Collagen triple helix repeat (20 copi
453	Collagen	9/11	376-430	41.4	2.4e-10	Collagen triple helix repeat (20 copi
453	Collagen	10/11	433-492	44.9	2.8e-11	Collagen triple helix repeat (20 copi
453	Collagen	11/11	495-535	30.3	2.2e-07	Collagen triple helix repeat (20 copi
453	C1q	1/1	576-700	263.0	4.8e-76	C1q domain
455	Transposase_22	1/1	2-28	11.7	0.0024	L1 transposable element
456	Ribosomal_S28e	1/1	57-97	41.9	5.5e-13	Ribosomal protein S28e
457	LRR	1/11	49-72	7.0	0.55	Leucine Rich Repeat
457	LRR	2/11	73-96	9.6	0.099	Leucine Rich Repeat
457	LRR	3/11	97-108	7.9	0.29	Leucine Rich Repeat
457	LRR	4/11	118-142	7.4	0.42	Leucine Rich Repeat
457	LRR	5/11	143-166	3.0	7.3	Leucine Rich Repeat
457	LRR	6/11	349-372	3.2	6.5	Leucine Rich Repeat
457	LRR	7/11	373-397	7.7	0.33	Leucine Rich Repeat
457	LRR	8/11	398-442	11.4	0.03	Leucine Rich Repeat
457	LRR	9/11	444-466	12.8	0.013	Leucine Rich Repeat
457	LRR	10/11	467-488	13.2	0.0098	Leucine Rich Repeat
457	LRR	11/11	489-512	0.4	39	Leucine Rich Repeat
457	LRRCT	1/1	550-575	18.2	8.9e-05	Leucine rich repeat C-terminal domain
457	TIR	1/1	636-774	113.0	9.6e-34	TIR domain
460	UPAR_LY6	1/1	23-101	30.8	3.9e-06	u-PAR/Ly-6 domain
460	Activin_recp	1/1	72-107	7.4	0.058	Activin types I and II receptor domain
461	UPAR_LY6	1/1	123-161	11.7	0.099	u-PAR/Ly-6 domain

TABLE 3B

SEQ ID	Model	Repeats	Position	Score	E_value	Description
462	Pep_M12B_propep	1/1	33-148	174.5	8.4e-56	Reprolysin family propeptide
462	Reprolysin	1/1	158-355	342.1	7.6e-100	Reprolysin (M12B) family zinc metallo
462	Disintegrin	1/2	373-384	8.2	0.029	Disintegrin
462	Disintegrin	2/2	413-483	91.9	1.6e-28	Disintegrin
462	EGF	1/2	489-508	2.5	7.5	EGF-like domain
462	EGF	2/2	625-649	11.5	0.024	EGF-like domain
462	SBP56	1/1	638-647	5.8	0.057	56kDa selenium binding protein (SBP56
463	Pep_M12B_propep	1/1	33-148	174.5	8.4e-56	Reprolysin family propeptide
463	Reprolysin	1/1	158-329	292.7	5.6e-85	Reprolysin (M12B) family zinc metallo
464	Reprolysin	1/1	41-72	21.2	1.8e-05	Reprolysin (M12B) family zinc metalloprot
464	Disintegrin	1/2	90-99	8.3	0.029	Disintegrin
464	Disintegrin	2/2	102-136	41.0	1.5e-12	Disintegrin
465	Pep_M12B_propep	1/1	1-83	113.2	4.3e-36	Reprolysin family propeptide
465	Reprolysin	1/1	93-107	18.7	8.6e-05	Reprolysin (M12B) family zinc metallo
465	Disintegrin	1/1	106-140	41.0	1.5e-12	Disintegrin
466	Duffy_binding	1/1	47-103	9.5	0.00084	Plasmodium Duffy binding protein
467	Duffy_binding	1/1	47-95	4.3	0.032	Plasmodium Duffy binding protein
468	Duffy_binding	1/1	47-103	9.5	0.00084	Plasmodium Duffy binding protein
469	Duffy_binding	1/1	33-80	7.4	0.0036	Plasmodium Duffy binding protein

TABLE 4

SEQ ID NO:	max S (Maximum Score)	mean S (Mean Score)	Position of Cleavage Site in Amino Acid Sequence
238	0.974	0.882	17
241	0.976	0.902	26
249	0.970	0.503	45
248	0.989	0.960	17
249	0.989	0.960	17
253	0.993	0.965	18
255	0.916	0.485	30
257	0.965	0.894	33
258	0.924	0.765	22
260	0.987	0.658	45
261	0.923	0.751	33
262	0.937	0.871	22
268	0.988	0.887	35
269	0.987	0.865	38
271	0.981	0.955	19
272	0.903	0.571	48
273	0.973	0.888	17
275	0.945	0.812	22
276	0.945	0.812	22
277	0.945	0.812	22
279	0.936	0.757	30
285	0.939	0.868	18
289	0.950	0.801	21
290	0.950	0.808	21
297	0.964	0.666	42
298	0.988	0.958	21
299	0.996	0.977	18
300	0.988	0.958	21
301	0.932	0.766	17
303	0.915	0.833	22
304	0.983	0.952	16
313	0.993	0.950	23
315	0.977	0.959	21
318	0.971	0.887	22
319	0.972	0.949	19
321	0.977	0.698	46
325	0.995	0.950	17
331	0.989	0.972	18
332	0.995	0.971	14
335	0.913	0.583	25
336	0.912	0.714	19
339	0.925	0.610	39
341	0.955	0.933	13
345	0.956	0.848	25
350	0.978	0.887	18
353	0.948	0.836	16
354	0.986	0.971	18
355	0.969	0.913	18
357	0.978	0.905	17
359	0.973	0.891	25
360	0.954	0.791	19
364	0.934	0.518	41
365	0.977	0.959	21
366	0.977	0.959	21

TABLE 4

SEQ ID NO:	max S (Maximum Score)	mean S (Mean Score)	Position of Cleavage Site in Amino Acid Sequence
367	0.882	0.622	16
368	0.979	0.938	17
369	0.978	0.842	20
370	0.960	0.809	31
371	0.964	0.790	16
375	0.944	0.809	20
376	0.896	0.771	13
379	0.939	0.523	19
380	0.948	0.855	17
386	0.908	0.583	45
387	0.895	0.527	26
388	0.963	0.889	23
394	0.980	0.906	25
397	0.934	0.784	24
400	0.963	0.844	28
401	0.963	0.844	28
402	0.987	0.924	24
409	0.933	0.713	30
415	0.984	0.923	20
416	0.957	0.886	19
417	0.972	0.727	20
418	0.890	0.534	22
419	0.926	0.704	34
420	0.923	0.602	23
421	0.966	0.833	20
422	0.969	0.880	16
423	0.951	0.814	26
424	0.971	0.882	24
426	0.957	0.894	18
427	0.936	0.649	19
428	0.980	0.871	23
429	0.949	0.806	18
431	0.888	0.724	14
432	0.979	0.926	22
433	0.907	0.651	23
434	0.989	0.832	36
437	0.921	0.692	34
439	0.957	0.874	28
440	0.957	0.874	28
441	0.939	0.749	32
442	0.985	0.896	22
443	0.993	0.916	20
444	0.993	0.916	20
445	0.970	0.851	37
446	0.973	0.829	30
447	0.944	0.710	26
451	0.974	0.920	22
453	0.990	0.920	28
454	0.984	0.746	26
456	0.979	0.890	26
460	0.985	0.898	22
461	0.996	0.691	49
462	0.955	0.933	13
463	0.955	0.933	13

TABLE 4

SEQ ID NO:	max S (Maximum Score)	mean S (Mean Score)	Position of Cleavage Site in Amino Acid Sequence
466	0.952	0.796	21
467	0.952	0.796	21
468	0.952	0.796	21
469	0.952	0.796	21
470	0.952	0.796	21

TABLE 5

SEQ ID	Accession No.	Genomic Location
1	gi15290868	11
2	gi10048073	2
3	gi9407868	1
4	gi28570413	11
5	gi18087689	8
6	gi11276216	17
7	gi9187146	1
8	gi20198695	11
9	gi8118474	11q24
10	gi19572477	1
11	gi21844559	5
12	gi17226706	17
13	gi13559997	9q31.2-32
14	gi16214577	9
15	gi16214577	9
16	gi13559997	9q31.2-32
17	gi3849820	17
18	gi19745067	17
19	gi19745067	17
20	gi15209407	10
21	gi21218133	18p
22	gi24110949	3p
23	gi10047952	2
24	gi20304074	2
25	gi5931541	22q11.2
26	gi2580478	9q34
27	gi7161187	1q23.1-24.1
28	gi7161187	1q23.1-24.1
29	gi7161187	1q23.1-24.1
30	gi7161187	1q23.1-24.1
31	gi7161187	1q23.1-24.1
32	gi7161187	1q23.1-24.1
33	gi15011674	15q21.3
34	gi8099866	15
35	gi10185444	9
36	gi17026193	14
37	gi15777898	14
38	gi13992803	7
39	gi16306514	2
40	gi13560103	20
41	gi13560103	20
42	gi13560103	20
43	gi6693602	21
44	gi22597601	8
45	gi17149791	7
46	gi17149791	7
47	gi4902689	22q13.31-13.33
48	gi3169112	6p22.1-22.3
49	gi27884942	15
50	gi18497186	4
51	gi15187252	16
52	gi24418064	8q24.2
53	gi8117631	11q24
54	gi8117631	11q24
55	gi8117631	11q24
56	gi27436841	17

TABLE 5

SEQ ID	Accession No.	Genomic Location
57	gi7705162	3
58	gi7960353	3
59	gi8119029	11q23
60	gi24418256	2
61	gi8122261	19
62	gi5441941	22q12.1-qter
63	gi10440757	2
64	gi10440757	2
65	gi10440757	2
66	gi25137136	1q23.1-24.1
67	gi22094313	19
68	gi8118827	11q22
69	gi21743744	19
70	gi17488717	8
71	gi17155015	16q24.3
72	gi12666964	6q23.1-24.1
77	gi18542958	16
78	gi24080647	8cen
79	gi6562059	22q13.1-13.32
80	gi16972764	1q25.1-31.3
83	gi11323318	20
84	gi17384050	10
85	gi10178737	1
87	gi13699261	12
88	gi28201743	15
89	gi14336615	11
90	gi3342735	19
91	gi32141371	16
92	gi18477278	9p34.1-35.1
94	gi13357313	8
95	gi13357313	8
96	gi13507299	9q33
97	gi18087658	21p11-q21.1
99	gi2076723	7q21
101	gi9663995	11q
102	gi4938290	1p35.1-36.12
103	gi13186087	14
104	gi14572559	9q34.11-34.3
106	gi16030143	9
107	gi9795658	7p22
108	gi21592159	8
109	gi5123976	4p16
110	gi7622528	12
111	gi29243343	11p
112	gi29243343	11p
113	gi17384056	9
114	gi17384056	9
115	gi6624940	20
116	gi8118732	18q11.2
117	gi11276211	17
118	gi4760420	4p16
119	gi5926691	6p21.3
120	gi8119068	18p11.3
121	gi8119068	18p11.3
123	gi1730464	11
124	gi15321567	2

TABLE 5

SEQ ID	Accession No.	Genomic Location
125	gi11493342	13
126	gi27902328	11
127	gi21623946	11
128	gi13094226	11q
129	gi19571564	9q22.2-31.1
130	gi16972764	1q25.1-31.3
131	gi16972764	1q25.1-31.3
132	gi19774339	10
133	gi6706820	6
134	gi12666277	10
135	gi18056702	2
136	gi21263225	19
137	gi17939962	11q
138	gi17425233	11q
139	gi13992781	2
140	gi17488656	8
141	gi8117363	18q23
142	gi13396423	13q33.1-34
143	gi13396423	13q33.1-34
144	gi17939979	11q
145	gi23396287	17
146	gi2342716	16
147	gi10803419	6p21.2-22.1
148	gi8119063	11q22
149	gi13273725	9
150	gi22208303	Xq12
151	gi18425273	5
152	gi13027555	17
153	gi15421899	17
154	gi29568034	19
155	gi29568034	19
156	gi29568034	19
157	gi2370075	Xq21.1
158	gi2370075	Xq21.1
160	gi21622769	11
161	gi20303530	10
162	gi29568034	19
163	gi29568034	19
164	gi29568034	19
165	gi13897297	14
166	gi13897297	14
167	gi14284833	14
168	gi15799575	19
169	gi28201476	19
170	gi16195220	19
171	gi10129456	9
172	gi10048054	10
173	gi5419768	6q12-13
175	gi13929477	16
176	gi21637457	5
177	gi5911819	22
178	gi5911819	22
179	gi20334304	18p
180	gi17149680	11
181	gi3132349	21q22.1
182	gi12584735	1

TABLE 5

SEQ ID	Accession No.	Genomic Location
183	gi7577612	12
184	gi23307998	19
185	gi7981303	20q13.2-13.33
186	gi22532577	11
187	gi32469520	p22-p21
188	gi15022678	16
189	gi7288048	20
190	gi27645810	9p13.1-13.3
191	gi10803419	6p21.2-22.1
192	gi22203176	3
193	gi10803524	10
194	gi13897270	14
195	gi11228439	Xq13
196	gi18072229	2
197	gi21206312	8
198	gi17154303	1
199	gi11071931	11q23
200	gi11119454	19
201	gi23307998	19
202	gi23307998	19
203	gi23307998	19
204	gi13751339	9
205	gi13751339	9
206	gi13751339	9
207	gi21206312	8
208	gi21206312	8
209	gi21206312	8
210	gi10047694	3
211	gi11276160	18
212	gi21747451	19
213	gi24270774	17
214	gi14718389	2
215	gi2734091	16
216	gi2734091	16
217	gi14190714	18
218	gi9801308	1p34.1-35.3
219	gi20303530	10
220	gi5042403	19
221	gi5042403	19
222	gi27877441	4
223	gi9588441	1p31.3-33
225	gi21206312	8
226	gi16944847	9
227	gi16030143	9
228	gi28557946	8
229	gi20330977	8
230	gi20330977	8
231	gi8117631	11q24
232	gi8117631	11q24
233	gi8117631	11q24
234	gi8117631	11q24
235	gi8117631	11q24

TABLE 6

SEQ ID	Number of Transmembrane Domains	Total Score	For each transmembrane domain, its amino acid position and its TM Pred Score
237	1	1587	673-689:1587
238	1	1758	216-234:1758
255	1	3055	457-473:3055
262	1	2929	227-250:2929
263	1	3217	455-472:3217
265	1	3310	455-476:3310
266	1	3217	455-472:3217
267	1	3217	469-486:3217
274	1	1715	41-67:1715
281	3	5496	70-85:2169 189-206:1230 231-252:2097
282	1	1470	51-66:1470
285	1	3083	195-213:3083
292	9	14963	66-81:1206 87-103:1480 289-307:1229 342-361:1910 911-930:1458 961-977:1485 999-1015:1929 1036-1048:1706 1070-1086:2560
294	1	1451	2173-2191:1451
297	2	3325	147-162:1306 254-273:2019
299	3	6916	272-288:2801 323-343:1575 626-646:2540
300	2	4034	104-120:1310 155-173:2724
304	1	2801	274-291:2801
307	6	10261	42-69:2558 78-102:1226 155-171:1769 205-221:2265 241-258:1201 291-306:1242
308	1	1853	66-81:1853
309	1	1853	66-81:1853
311	2	3054	65-81:1510 154-174:1544
312	1	1360	668-690:1360
316	1	3116	406-427:3116
317	3	5762	64-79:1689 124-141:1689 184-205:2384
321	1	1738	295-310:1738
324	2	4456	66-82:2701 110-126:1755
328	2	2797	78-91:1294 113-128:1503
334	5	8786	214-232:1714 261-286:2149 359-376:2223 393-415:1222 426-447:1478
336	1	1619	728-744:1619
337	8	15409	76-92:1313 101-121:2655 134-151:1463 176-194:2732 202-217:1469 240-257:1784 278-293:1206 303-320:2787
338	1	1762	1042-1060:1762
339	3	4501	265-280:1571 342-358:1659 404-421:1271
340	6	11252	153-168:1912 182-199:2580 220-236:1334 248-263:2458 279-294:1217 311-329:1751
341	1	1472	654-674:1472
344	2	3175	484-503:1730 613-632:1445
345	1	2846	225-247:2846
346	1	1459	172-189:1459
347	1	1459	381-398:1459
352	2	2764	772-791:1300 2058-2074:1464
354	1	2180	110-124:2180
355	5	10377	71-90:1665 154-171:1865 185-199:1592 233-255:2690 300-314:2565
356	5	10377	76-95:1665 159-176:1865 190-204:1592 238-260:2690 305-319:2565
363	1	2897	207-233:2897
368	1	3181	162-180:3181

TABLE 6

SEQ ID	Number of Transmembrane Domains	Total Score	For each transmembrane domain, its amino acid position and its TM Pred Score
371	1	3219	136-157:3219
377	1	2746	61-76:2746
389	1	2559	155-178:2559
390	1	2559	176-199:2559
392	4	7798	53-76:2169 167-184:1878 221-236:1565 261-278:2186
393	3	5629	150-167:1878 204-219:1565 244-261:2186
394	1	3033	95-118:3033
400	1	1341	713-729:1341
404	1	1512	95-109:1512
405	1	2976	67-84:2976
415	1	2904	217-236:2904
421	1	1533	163-181:1533
425	1	2350	129-149:2350
430	1	1373	56-77:1373
443	1	2065	109-125:2065
446	1	3354	420-442:3354
450	1	1335	131-147:1335
451	1	1335	292-308:1335
457	1	2970	578-596:2970
462	1	1472	686-706:1472

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
1	236	510	850	784_7010
1	236	644	984	790_14876
1	236	748	1088	810_7
2	237	611	951	788_7192
3	238	707	1047	802_99
3	238	708	1048	802_100
3	238	768	1108	815_1
4	239			
5	240	720	1060	803_830
5	240	744	1084	808_111
6	241	548	888	784_9546
6	241	690	1030	795_96
6	241	783	1123	819_4
7	242	751	1091	811_2
7	242	757	1097	814_1
7	242	784	1124	819_36
8	243	691	1031	795_301
8	243	692	1032	795_302
8	243	775	1115	816_14
9	244	752	1092	811_23
10	245	482	822	784_3067
11	246	495	835	784_5316
11	246	696	1036	796_121
11	246	698	1038	797_121
12	247			
13	248	490	830	784_4111
13	248	701	1041	799_46
13	248	712	1052	803_34
14	249	480	820	784_2832
14	249	712	1052	803_34
15	250	492	832	784_4671
15	250	712	1052	803_34
16	251	490	830	784_4111
16	251	701	1041	799_46
16	251	712	1052	803_34
17	252	536	876	784_8254
17	252	559	899	785_2244
18	253	726	1066	805_203
19	254	726	1066	805_203
20	255	483	823	784_3137
21	256	513	853	784_7230
21	256	749	1089	810_227
22	257	514	854	784_7233
23	258	623	963	790_1871
23	258	625	965	790_3086
23	258	717	1057	803_547
24	259	535	875	784_8246
24	259	608	948	787_10343
25	260	697	1037	796_144

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
25	260	699	1039	797_144
25	260	760	1100	814_28
26	261	551	891	785_155
26	261	694	1034	795_482
26	261	702	1042	799_60
27	262	503	843	784_6724
27	262	638	978	790_11759
27	262	689	1029	794_321
28	263	756	1096	813_301
29	264	756	1096	813_301
30	265	756	1096	813_301
31	266	756	1096	813_301
32	267	765	1096	813_301
33	268	642	982	790_14016
33	268	785	1125	819_125
34	269	540	880	784_8624
34	269	761	1101	814_32
35	270	500	840	784_5987
36	271	500	840	784_5987
37	272	522	862	784_7453
38	273	539	879	784_8622
38	273	588	928	787_7723
39	274			
40	275	572	912	787_2647
41	276	567	907	787_2006
41	276	572	912	787_2647
42	277	572	912	787_2647
43	278	479	819	784_2681
43	278	486	826	784_3464
43	278	575	915	787_4441
44	279	471	811	784_429
44	279	497	837	784_5476
44	279	776	1116	816_43
45	280	489	829	784_4040
45	280	506	846	784_6870
45	280	665	1005	791_1838
46	281	489	829	784_4040
46	281	506	846	784_6870
46	281	665	1005	791_1838
47	282	677	1017	792_3878
48	283	484	824	784_3248
48	283	610	950	787_10389
49	284	568	908	787_2040
49	284	579	919	787_5487
49	284	740	1080	806_1017
50	285	538	878	784_8515
50	285	560	900	785_2334
51	286	473	813	784_875
51	286	687	1027	792_7767

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
51	286	786	1126	819_179
52	287	516	856	784_7273
53	288	686	1026	792_7097
53	288	727	1067	806_68
54	289	686	1026	792_7097
54	289	727	1067	806_68
55	290	686	1026	792_7097
55	290	727	1067	806_68
56	291	682	1022	792_4929
56	291	777	1117	816_49
57	292	501	841	784_6261
57	292	584	924	787_6675
58	293	565	905	787_123
59	294	576	916	787_4535
59	294	646	986	790_17432
59	294	647	987	790_17433
60	295	541	881	784_8636
60	295	787	1127	819_189
61	296	523	863	784_7497
62	297	574	914	787_4251
62	297	577	917	787_4937
63	298	606	946	787_9212
63	298	710	1050	802_339
63	298	788	1128	819_193
64	299	606	946	787_9212
64	299	710	1050	802_339
64	299	788	1128	819_193
65	300	606	946	787_9212
65	300	710	1050	802_339
65	300	788	1128	819_193
66	301	519	859	784_7361
66	301	652	992	790_20838
66	301	675	1015	792_3608
67	302			
68	303	789	1129	819_194
68	303	790	1130	819_195
68	303	791	1131	819_196
69	304	561	901	785_2811
69	304	769	1109	815_22
70	305			
71	306	633	973	790_8459
71	306	657	997	790_24619
71	306	658	998	790_24626
72	307	496	836	784_5458
72	307	570	910	787_2123
73	308	629	969	790_6152
73	308	713	1053	803_132
74	309	629	969	790_6152
74	309	713	1053	803_132

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
75	310	629	969	790_6152
75	310	713	1053	803_132
76	311	629	969	790_6152
76	311	713	1053	803_132
77	312	502	842	784_6478
77	312	614	954	789_872
77	312	648	988	790_18038
78	313	793	1133	819_224
78	313	801	1141	819_417
78	313	802	1142	819_418
79	314	530	870	784_7932
79	314	591	931	787_7886
80	315	562	902	785_2845
80	315	803	1143	819_421
81	316	498	838	784_5730
82	317	524	864	784_7600
82	317	609	949	787_10362
83	318	550	890	784_10222
83	318	634	974	790_8816
83	318	728	1068	806_143
84	319	546	886	784_9103
85	320	512	852	784_7225
85	320	703	1043	799_72
85	320	779	1119	816_72
86	321	529	869	784_7782
87	322	651	991	790_19661
88	323	593	933	787_7964
89	324	671	1011	792_2342
89	324	755	1095	812_111
90	325	508	848	784_6946
90	325	809	1149	819_678
91	326	594	934	787_7980
92	327	493	833	784_4821
92	327	781	1121	816_196
92	327	794	1134	819_273
93	328	520	860	784_7366
93	328	596	936	787_8036
94	329	597	937	787_8045
94	329	655	995	790_23678
95	330	597	937	787_8045
96	331	525	865	784_7634
96	331	526	866	784_7655
96	331	598	938	787_8052
97	332	661	1001	790_27622
97	332	683	1023	792_6308
98	333			
99	334	620	960	790_105
99	334	640	980	790_12371
99	334	688	1028	793_94

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
100	335	528	868	784_7755
100	335	796	1136	819_302
101	336	599	939	787_8109
101	336	626	966	790_3197
102	337	527	867	784_7658
102	337	600	940	787_8111
103	338	770	1110	815_41
103	338	797	1137	819_308
104	339	587	927	787_7662
104	339	746	1086	809_213
105	340	580	920	787_5697
105	340	664	1004	791_577
106	341			
107	342	515	855	784_7261
107	342	780	1120	816_98
108	343	476	816	784_2188
109	344			
110	345	676	1016	792_3857
110	345	798	1138	819_343
111	346	581	921	787_6059
111	346	674	1014	792_3539
111	346	725	1065	805_68
112	347	581	921	787_6059
112	347	674	1014	792_3539
112	347	725	1065	805_68
113	348	743	1083	808_79
114	349	743	1083	808_79
115	350	799	1139	819_373
115	350	800	1140	819_375
116	351	555	895	785_888
117	352	533	873	784_8131
117	352	601	941	787_8227
118	353	704	1044	799_85
118	353	762	1102	814_70
119	354	589	929	787_7763
119	354	693	1033	795_316
120	355	624	964	790_2755
120	355	714	1054	803_432
120	355	771	1111	815_59
121	356	624	964	790_2755
121	356	714	1054	803_432
121	356	771	1111	815_59
122	357	810	1150	819_682
123	358			
124	359	557	897	785_1597
125	360	566	906	787_181
125	360	778	1118	816_56
126	361	766	1106	814_164
127	362	509	849	784_6962

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
127	362	767	1107	814_167
128	363	521	861	784_7400
128	363	670	1010	792_1669
128	363	700	1040	799_20
129	364	729	1069	806_353
130	365	562	902	785_2845
130	365	803	1143	819_421
131	366	562	902	785_2845
131	366	803	1143	819_421
132	367			
133	368	632	972	790_8424
133	368	711	1051	802_425
133	368	772	1112	815_65
134	369	745	1085	809_50
135	370	478	818	784_2432
135	370	722	1062	804_308
136	371	563	903	785_2878
136	371	604	944	787_8798
137	372			
138	373			
139	374	532	872	784_8116
140	375	537	877	784_8471
141	376	553	893	785_765
141	376	558	898	785_2024
141	376	695	1035	796_28
142	377	773	1113	815_73
142	377	782	1122	818_60
143	378	773	1113	815_73
144	379	554	894	785_845
144	379	731	1071	806_423
145	380	732	1072	806_424
145	380	804	1144	819_454
146	381	586	926	787_7005
147	382	617	957	789_3980
148	383	662	1002	790_27696
148	383	774	1114	815_141
148	383	805	1145	819_468
149	384	488	828	784_3985
149	384	715	1055	803_534
149	384	716	1056	803_535
150	385	504	844	784_6798
150	385	750	1090	810_685
151	386	733	1073	806_456
151	386	806	1146	819_480
152	387	518	858	784_7301
153	388	613	953	788_13842
154	389	705	1045	802_53
155	390	705	1045	802_53
156	391	705	1045	802_53

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
157	392	754	1094	812_108
158	393	754	1094	812_108
159	394	627	967	790_5231
159	394	628	968	790_5232
159	394	741	1081	807_138
160	395	549	889	784_10220
160	395	607	947	787_9766
161	396	531	871	784_8001
161	396	603	943	787_8771
162	397	569	909	787_2097
162	397	615	955	789_1430
162	397	742	1082	808_62
163	398	569	909	787_2097
163	398	635	975	790_9670
163	398	742	1082	808_62
164	399	569	909	787_2097
164	399	635	975	790_9670
164	399	742	1082	808_62
165	400	474	814	784_1062
165	400	763	1103	814_112
166	401	730	1070	806_355
167	402	544	884	784_9018
167	402	795	1135	819_278
168	403	630	970	790_7151
168	403	656	996	790_24492
169	404	582	922	787_6147
169	404	631	971	790_7977
170	405	612	952	788_12683
171	406	505	845	784_6859
172	407	616	956	789_3199
173	408	605	945	787_8852
174	409	499	839	784_5939
175	410	618	958	789_5315
175	410	659	999	790_25550
175	410	721	1061	803_922
176	411	481	821	784_2986
177	412	758	1098	814_9
177	412	759	1099	814_10
178	413	758	1098	814_9
178	413	759	1099	814_10
179	414	764	1104	814_118
180	415	807	1147	819_574
181	416	592	932	787_7895
181	416	621	961	790_582
181	416	622	962	790_584
182	417	734	1074	806_694
182	417	753	1093	811_85
183	418	667	1007	791_3897
183	418	735	1075	806_697

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
185	419	650	990	790_18620
185	419	669	1009	792_66
185	419	685	1025	792_7077
185	420	723	1063	804_436
185	420	724	1064	804_437
185	420	765	1105	814_119
186	421			
187	422	564	904	785_2998
187	422	736	1076	806_734
188	423	485	825	784_3419
188	423	639	979	790_12222
188	423	663	1003	790_27760
189	424	543	883	784_8768
189	424	709	1049	802_227
189	424	792	1132	819_207
190	425	578	918	787_5204
190	425	747	1087	809_262
191	426	534	874	784_8214
192	427	645	985	790_16803
193	428	507	847	784_6881
193	428	738	1078	806_850
194	429	487	827	784_3632
194	429	585	925	787_6957
195	430			
196	431	602	942	787_8335
197	432	739	1079	806_871
198	433			
199	434	641	981	790_13752
199	434	672	1012	792_3125
199	434	673	1013	792_3131
200	435	472	812	784_824
200	435	475	815	784_1142
200	435	552	892	785_248
201	436	678	1018	792_3972
201	436	680	1020	792_3974
201	436	681	1021	792_3979
202	437	653	993	790_21179
202	437	668	1008	792_60
202	437	679	1019	792_3973
203	438	511	851	784_7113
203	438	649	989	790_18618
203	438	684	1024	792_7076
204	439	477	817	784_2330
204	439	808	1148	819_640
205	440	477	817	784_2330
205	440	571	911	787_2281
205	440	573	913	787_2967
206	441	477	817	784_2330
206	441	571	911	787_2281

TABLE 7

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.)*
206	441	573	913	787_2967
207	442			
208	443	542	882	784_8671
209	444	542	882	784_8671
210	445	595	935	787_8030
210	445	619	959	790_21
210	445	737	1077	806_828
211	446	494	834	784_5131
211	446	547	887	784_9193
211	446	718	1058	803_579
212	447			
213	448	556	896	785_1513
214	448	654	994	790_22798
214	449			
215	450			
216	451			
217	452	517	857	784_7275
217	452	590	930	787_7810
218	453	583	923	787_6566
219	454	719	1059	803_796
220	455	706	1046	802_64
221	456			
222	457	491	831	784_4613
222	457	545	885	784_9044
223	458			
224	459			
225	460			
226	461			
227	462			
228	463			
229	464			
230	465			
231	466	643	983	790_14421
231	466	660	1000	790_26186
231	466	666	1006	791_2167
232	467	666	1006	791_2167
233	468	636	976	790_11429
233	468	637	977	790_11454
233	468	666	1006	791_2167
234	469	636	976	790_11429
234	469	637	977	790_11454
234	469	666	1006	791_2167
235	470	666	1006	791_2167

*784_XXX = SEQ ID NO: XXX of Attorney Docket No. 784, U.S. Serial No. 09/488,725 filed 01/21/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 784CIP, U.S. Application Serial No. 09/552,317, filed April 25, 2000, which in turn is a parent application of

continuation-in-part application bearing Attorney Docket No. 784CIP3A/PCT, PCT Serial No. PCT/US00/35017 filed December 22, 2000, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

785_XXX = SEQ ID NO: XXX of Attorney Docket No. 785, U.S. Serial No. 09/491,404 filed 01/25/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 785CIP3/PCT, PCT Serial No. PCT/US01/02623 filed January 25, 2001, which is incorporated herein by reference in its entirety, including Tables, and Sequence Listing.

787_XXX = SEQ ID NO: XXX of Attorney Docket No. 787, U.S. Serial No. 09/496,914 filed 02/03/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 787CIP, U.S. Application Serial No. 09/560,875, filed April 27, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 787CIP3/PCT, PCT Serial No. PCT/US01/03800 filed February 5, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

788_XXX = SEQ ID NO: XXX of Attorney Docket No. 788, U.S. Serial No. 09/515,126 filed 02/28/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 788CIP, U.S. Application Serial No. 09/577,409, filed May 18, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 788CIP3/PCT, PCT Serial No. PCT/US01/04927 filed February 26, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

789_XXX = SEQ ID NO: XXX of Attorney Docket No. 789, U.S. Serial No. 09/519,705 filed 03/07/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 789CIP, U.S. Application Serial No. 09/574,454, filed May 19, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 789CIP3/PCT, PCT Serial No. PCT/US01/04941 filed March 5, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

790_XXX = SEQ ID NO: XXX of Attorney Docket No. 790, U.S. Serial No. 09/540,217 filed 03/31/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 790CIP, U.S. Application Serial No. 09/649,167, filed August 23, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 790CIP3/PCT, PCT Serial No. PCT/US01/08631 filed March 30, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

791_XXX = SEQ ID NO: XXX of Attorney Docket No. 791, U.S. Serial No. 09/552,929 filed 04/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 791CIP, U.S. Application Serial No. 09/770,160, filed January 26, 2001, which in turn is a parent application of

continuation-in-part application bearing Attorney Docket No. 791CIP3/PCT, PCT Serial No. PCT/US01/8656 filed April18, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

792_XXX = SEQ ID NO: XXX of Attorney Docket No. 792, U.S. Serial No. 09/577,408 filed May 18, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 792CIP3/PCT, PCT Serial No. PCT/US01/14827 filed May16, 2001, which is incorporated herein by reference in its entirety, including Tables, and Sequence Listing.

793_XXX = SEQ ID NO: XXX of Attorney Docket No. 793, U.S. Serial No. 09/654,935, filed September 01, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 793CIP/PCT, PCT Serial No. PCT/US01/27093, filed August 31, 2001, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

794_XXX = SEQ ID NO: XXX of Attorney Docket No. 794, U.S. Serial No. 09/659,671, filed September 11, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 794CIP/PCT, PCT Serial No. PCT/US01/26015 filed September 10, 2001, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

795_XXX = SEQ ID NO: XXX of Attorney Docket No. 795, U.S. Serial No. 09/687,527 filed October 12, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 795CIP/PCT, PCT Serial No. PCT/US01/27760 filed October 11, 2001, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

796_XXX = SEQ ID NO: XXX of Attorney Docket No. 796, U.S. Serial No. 09/707,351 filed November 06, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 796/785CIP/PCT, PCT Serial No. PCT/US01/02723 filed January 25, 2001, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

797_XXX = SEQ ID NO: XXX of Attorney Docket No. 797, U.S. Serial No. 09/714,936 filed November 17, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 797CIP/PCT, PCT Serial No. PCT/US01/42950 filed November 16, 2001, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

799_XXX = SEQ ID NO: XXX of Attorney Docket No. 799, U.S. Serial No. 09/728,952 filed November 30, 2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 799CIP/PCT, PCT Serial No. PCT/US01/47004 filed November 30, 2001, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

802_XXX = SEQ ID NO: XXX of Attorney Docket No. 802, U.S. Serial No. 09/774,528 filed January 30, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This

application is the parent application of a continuation-in-part application bearing Attorney Docket No. 802CIP/PCT, PCT Serial No. PCT/US02/01222 filed January 29, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

803_XXX = SEQ ID NO: XXX of Attorney Docket No. 803, U.S. Serial No. 09/799,451 filed March 05, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 803CIP/PCT, PCT Serial No. PCT/US02/05095 filed March 05, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

804_XXX = SEQ ID NO: XXX of Attorney Docket No. 804, U.S. Serial No. 09/810,173 filed March 15, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 804CIP/PCT, PCT Serial No. PCT/US02/05109 filed March 14, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

805_XXX = SEQ ID NO: XXX of Attorney Docket No. 805, U.S. Provisional Serial No. 60/306,971 filed July 21, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 805A, U.S. Serial No. 10/112,944 filed March 28, 2002, which is the parent application of a continuation-in-part application bearing Attorney Docket No. 805A/PCT, PCT Serial No. PCT/US02/22858 filed July 19, 2002, both of which are incorporated herein by reference in their entirety, including Tables and Sequence Listing.

806_XXX = SEQ ID NO: XXX of Attorney Docket No. 806, U.S. Provisional Serial No. 60/311,261 filed August 09, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 806A, U.S. Serial No. 10/219,382 filed August 09, 2002, which is the parent application of a continuation-in-part application bearing Attorney Docket No. 806CIP/PCT, PCT Serial No. PCT/US02/25485 filed August 09, 2002, both of which are incorporated herein by reference in their entirety, including Tables and Sequence Listing.

807_XXX = SEQ ID NO: XXX of Attorney Docket No. 807, U.S. Provisional Serial No. 60/322,511 filed September 13, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 807A, U.S. Serial No. 10/243,552 filed September 12, 2002, which is the parent application of a continuation-in-part application bearing Attorney Docket No. 807ACIP/PCT, PCT Serial No. PCT/US02/29001 filed September 13, 2002, both of which are incorporated herein by reference in their entirety, including Tables and Sequence Listing.

808_XXX = SEQ ID NO: XXX of Attorney Docket No. 808, U.S. Provisional Serial No. 60/323,349 filed September 18, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 808A, U.S. Serial No. 10/245,817 filed September 16, 2002, which is the parent application of a continuation-in-part application bearing Attorney Docket No. 808ACIP/PCT, PCT Serial No.

PCT/US02/29636 filed September 18, 2002, both of which are incorporated herein by reference in their entirety, including Tables and Sequence Listing.

809_XXX = SEQ ID NO: XXX of Attorney Docket No. 809, U.S. Provisional Serial No. 60/323,739 filed September 19, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 809A, U.S. Serial No. 10/245,014 filed September 16, 2002, which is the parent application of a continuation-in-part application bearing Attorney Docket No. 809ACIP/PCT, PCT Serial No. PCT/US02/29964 filed September 19, 2002, both of which are incorporated herein by reference in their entirety, including Tables and Sequence Listing.

810_XXX = SEQ ID NO: XXX of Attorney Docket No. 810, U.S. Provisional Serial No. 60/324,631 filed September 24, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 810CIP/PCT, PCT Serial No. PCT/US02/30474 filed September 24, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

811_XXX = SEQ ID NO: XXX of Attorney Docket No. 811, U.S. Provisional Serial No. 60/339,739 filed December 10, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

812_XXX = SEQ ID NO: XXX of Attorney Docket No. 812, U.S. Provisional Serial No. 60/339,453 filed December 11, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 812A, U.S. Serial No. 10/128,558, which is the parent application of a continuation-in-part application bearing Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, both of which are incorporated herein by reference in their entirety, including Tables and Sequence Listing.

813_XXX = SEQ ID NO: XXX of Attorney Docket No. 813, U.S. Provisional Serial No. 60/340,187 filed December 12, 2001, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

814_XXX = SEQ ID NO: XXX of Attorney Docket No. 814, U.S. Provisional Serial No. 60/365,384 filed March 14, 2002, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

815_XXX = SEQ ID NO: XXX of Attorney Docket No. 815, U.S. Provisional Serial No. 60/365,091 filed March 14, 2002, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing

TABLE 7

Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

816_XXX = SEQ ID NO: XXX of Attorney Docket No. 816, U.S. Provisional Serial No. 60/365,264 filed March 14, 2002, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

818_XXX = SEQ ID NO: XXX of Attorney Docket No. 818, U.S. Provisional Serial No. 60/372,381 filed April 12, 2002, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the provisional application to which priority is claimed in the utility application bearing Attorney Docket No. 820/PCT, PCT Serial No. PCT/US02/39555 filed December 10, 2002, which is incorporated herein by reference in its entirety, including Tables and Sequence Listing.

819_XXX = SEQ ID NO: XXX of Attorney Docket No. 819, U.S. Provisional Serial No. 60/416,186 filed October 02, 2002, the entire disclosure of which, including Tables and Sequence Listing, is incorporated herein by reference in its entirety.

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120
121	121
122	122
123	123
124	124
125	125
126	126
127	127
128	128
129	129
130	130
131	131
132	132
133	133
134	134
135	135
136	136
137	137
138	138
139	139
140	140
141	141
142	142
143	143
144	144
145	145
146	146
147	147
148	148
149	149
150	150
151	151
152	152
153	153
154	154
155	155
156	156
157	157
158	158
159	159
160	160
161	161
162	162
163	163
164	164
165	165

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
166	166
167	167
168	168
169	169
170	170
171	171
172	172
173	173
174	174
175	175
176	176
177	178
178	179
179	180
180	181
181	182
182	183
183	184
184	185
185	186
186	187
187	188
188	189
189	190
190	191
191	192
192	193
193	194
194	195
195	196
196	197
197	198
198	199
199	200
200	201
201	202
202	203
203	204
204	205
205	206
206	207
207	208
208	209
209	210
210	211
211	212
212	213
213	214
214	215
215	216
216	217
217	218
218	219
219	220
220	221

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
221	222
222	223
223	224
224	225
225	226
226	227
227	228
228	229
229	230
230	231
231	232
232	233
233	234
234	235
235	236
236	237
237	238
238	239
239	240
240	241
241	242
242	243
243	244
244	245
245	246
246	247
247	248
248	249
249	250
250	251
251	252
252	253
253	254
254	255
255	256
256	257
257	258
258	259
259	260
260	261
261	262
262	263
263	264
264	265
265	266
266	267
267	268
268	269
269	270
270	271
271	272
272	273
273	274
274	275
275	276

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
276	277
277	278
278	279
279	280
280	281
281	282
282	283
283	284
284	285
285	286
286	287
287	288
288	289
289	290
290	291
291	292
292	293
293	294
294	295
295	296
296	297
297	298
298	299
299	300
300	301
301	302
302	303
303	304
304	305
305	306
306	307
307	308
308	309
309	310
310	311
311	312
312	313
313	314
314	315
315	316
316	317
317	318
318	319
319	320
320	321
321	322
322	323
323	324
324	325
325	326
326	327
327	328
328	329
329	330
330	331

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
331	332
332	333
333	334
334	335
335	336
336	337
337	338
338	339
339	340
340	341
341	342
342	343
343	344
344	345
345	346
346	347
347	348
348	349
349	350
350	351
351	352
352	353
353	354
354	355
355	356
356	357
357	358
358	359
359	360
360	361
361	362
362	363
363	364
364	365
365	366
366	367
367	368
368	369
369	370
370	371
371	372
372	373
373	374
374	375
375	376
376	377
377	378
378	379
379	380
380	381
381	382
382	383
383	384
384	385
385	386

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
386	387
387	388
388	389
389	390
390	391
391	392
392	393
393	394
394	395
395	396
396	397
397	398
398	399
399	400
400	401
401	402
402	403
403	404
404	405
405	406
406	407
407	408
408	409
409	410
410	411
411	412
412	414
413	415
414	416
415	417
416	418
417	419
418	420
419	421
420	422
421	423
422	424
423	425
424	426
425	427
426	428
427	429
428	430
429	431
430	432
431	433
432	434
433	435
434	436
435	437
436	438
437	439
438	440
439	441
440	442

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
441	443
442	444
443	445
444	446
445	447
446	448
447	449
448	450
449	451
450	452
451	453
452	454
453	455
454	456
455	457
456	458
457	459
458	460
459	461
460	462
461	463
462	464
463	465
464	466
465	467
466	468
467	469
468	470
469	471
470	472
471	473
472	474
473	475
474	476
475	477
476	478
477	479
478	480
479	481
480	482
481	483
482	484
483	485
484	486
485	487
486	488
487	489
488	490
489	491
490	492
491	493
492	494
493	495
494	496
495	497

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
496	498
497	499
498	500
499	501
500	502
501	503
502	504
503	505
504	506
505	507
506	508
507	509
508	510
509	511
510	512
511	513
512	514
513	515
514	516
515	517
516	518
517	519
518	520
519	521
520	522
521	523
522	524
523	525
524	526
525	527
526	528
527	529
528	530
529	531
530	532
531	533
532	534
533	535
534	536
535	537
536	538
537	539
538	540
539	541
540	542
541	543
542	544
543	545
544	546
545	547
546	548
547	549
548	550
549	551
550	552

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
551	554
552	555
553	557
554	558
555	559
556	560
557	561
558	562
559	563
560	564
561	565
562	566
563	567
564	568
565	569
566	570
567	571
568	572
569	573
570	574
571	575
572	576
573	577
574	578
575	579
576	580
577	581
578	582
579	583
580	584
581	585
582	586
583	587
584	588
585	589
586	590
587	591
588	592
589	593
590	594
591	595
592	596
593	597
594	598
595	599
596	600
597	601
598	602
599	603
600	604
601	605
602	606
603	607
604	608
605	609

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
606	610
607	611
608	612
609	613
610	614
611	615
612	616
613	617
614	618
615	619
616	620
617	621
618	622
619	623
620	624
621	625
622	626
623	627
624	628
625	629
626	630
627	631
628	632
629	633
630	634
631	635
632	636
633	637
634	638
635	639
636	640
637	641
638	642
639	643
640	644
641	645
642	646
643	647
644	648
645	649
646	650
647	651
648	652
649	653
650	654
651	655
652	656
653	657
654	658
655	659
656	660
657	661
658	662
659	663
660	664

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
661	665
662	666
663	667
664	668
665	669
666	670
667	671
668	672
669	673
670	674
671	675
672	676
673	677
674	678
675	679
676	680
677	681
678	682
679	683
680	684
681	685
682	686
683	687
684	688
685	689
686	690
687	691
688	692
689	693
690	694
691	695
692	696
693	697
694	698
695	699
696	700
697	701
698	702
699	703
700	704
701	705
702	706
703	707
704	708
705	709
706	710
707	711
708	712
709	713
710	714
711	715
712	716
713	717
714	718
715	719

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
716	720
717	721
718	722
719	723
720	724
721	725
722	726
723	727
724	728
725	729
726	730
727	731
728	732
729	733
730	734
731	735
732	736
733	737
734	739
735	740
736	741
737	742
738	743
739	744
740	745
741	746
742	747
743	748
744	749
745	750
746	751
747	752
748	753
749	754
750	755
751	756
752	757
753	758
754	759
755	760
756	761
757	762
758	763
759	764
760	765
761	766
762	767
763	768
764	769
765	770
766	771
767	772
768	773
769	774
770	775

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
771	776
772	777
773	778
774	779
775	780
776	781
777	782
778	783
779	784
780	785
781	786
782	787
783	788
784	789
785	790
786	791
787	792
788	793
789	794
790	795
791	796
792	797
793	798
794	799
795	800
796	801
797	802
798	803
799	804
800	805
801	806
802	807
803	808
804	809
805	810
806	811
807	812
808	813
809	814
810	815
811	816
812	817
813	818
814	819
815	820
816	821
817	822
818	823
819	824
820	825
821	826
822	827
823	828
824	829
825	830

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
826	831
827	832
828	833
829	834
830	835
831	836
832	837
833	838
834	839
835	840
836	841
837	842
838	843
839	844
840	845
841	846
842	847
843	848
844	849
845	850
846	851
847	852
848	853
849	854
850	855
851	856
852	857
853	858
854	859
855	860
856	861
857	862
858	863
859	864
860	865
861	866
862	867
863	868
864	869
865	870
866	871
867	872
868	873
869	874
870	875
871	876
872	877
873	878
874	879
875	880
876	881
877	882
878	883
879	884
880	885

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
881	886
882	887
883	888
884	889
885	890
886	891
887	892
888	893
889	894
890	895
891	897
892	898
893	900
894	901
895	902
896	903
897	904
898	905
899	906
900	907
901	908
902	909
903	910
904	911
905	912
906	913
907	914
908	915
909	916
910	917
911	918
912	919
913	920
914	921
915	922
916	923
917	924
918	925
919	926
920	927
921	928
922	929
923	930
924	931
925	932
926	933
927	934
928	935
929	936
930	937
931	938
932	939
933	940
934	941
935	942

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
936	943
937	944
938	945
939	946
940	947
941	948
942	949
943	950
944	951
945	952
946	953
947	954
948	955
949	956
950	957
951	958
952	959
953	960
954	961
955	962
956	963
957	964
958	965
959	966
960	967
961	968
962	969
963	970
964	971
965	972
966	973
967	974
968	975
969	976
970	977
971	978
972	979
973	980
974	981
975	982
976	983
977	984
978	985
979	986
980	987
981	988
982	989
983	990
984	991
985	992
986	993
987	994
988	995
989	996
990	997

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
991	998
992	999
993	1000
994	1001
995	1002
996	1003
997	1004
998	1005
999	1006
1000	1007
1001	1008
1002	1009
1003	1010
1004	1011
1005	1012
1006	1013
1007	1014
1008	1015
1009	1016
1010	1017
1011	1018
1012	1019
1013	1020
1014	1021
1015	1022
1016	1023
1017	1024
1018	1025
1019	1026
1020	1027
1021	1028
1022	1029
1023	1030
1024	1031
1025	1032
1026	1033
1027	1034
1028	1035
1029	1036
1030	1037
1031	1038
1032	1039
1033	1040
1034	1041
1035	1042
1036	1043
1037	1044
1038	1045
1039	1046
1040	1047
1041	1048
1042	1049
1043	1050
1044	1051
1045	1052

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
1046	1053
1047	1054
1048	1055
1049	1056
1050	1057
1051	1058
1052	1059
1053	1060
1054	1061
1055	1062
1056	1063
1057	1064
1058	1065
1059	1066
1060	1067
1061	1068
1062	1069
1063	1070
1064	1071
1065	1072
1066	1073
1067	1074
1068	1075
1069	1076
1070	1077
1071	1078
1072	1079
1073	1080
1074	1082
1075	1083
1076	1084
1077	1085
1078	1086
1079	1087
1080	1088
1081	1089
1082	1090
1083	1091
1084	1092
1085	1093
1086	1094
1087	1095
1088	1096
1089	1097
1090	1098
1091	1099
1092	1100
1093	1101
1094	1102
1095	1103
1096	1104
1097	1105
1098	1106
1099	1107
1100	1108

TABLE 8

SEQ ID NO:	SEQ ID NO: in Priority Application 60/458,824
1101	1109
1102	1110
1103	1111
1104	1112
1105	1113
1106	1114
1107	1115
1108	1116
1109	1117
1110	1118
1111	1119
1112	1120
1113	1121
1114	1122
1115	1123
1116	1124
1117	1125
1118	1126
1119	1127
1120	1128
1121	1129
1122	1130
1123	1131
1124	1132
1125	1133
1126	1134
1127	1135
1128	1136
1129	1137
1130	1138
1131	1139
1132	1140
1133	1141
1134	1142
1135	1143
1136	1144
1137	1145
1138	1146
1139	1147
1140	1148
1141	1149
1142	1150
1143	1151
1144	1152
1145	1153
1146	1154
1147	1155
1148	1156
1149	1157
1150	1158

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-235.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 99% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-235.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of any of the polynucleotides from SEQ ID NO: 1-235, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO: 236-470.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising of at least one of SEQ ID NO: 1-235.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.